Journal of Multivariate Analysis 101 (2010) 1656–1680

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Simultaneous confidence band and hypothesis test in generalised varying-coefficient models

Wenyang Zhang [a,*], Heng Peng [b]

[a] *Department of Mathematical Sciences, University of Bath, UK*

[b] *Department of Mathematics, Hong Kong Baptist University, Hong Kong*

ARTICLE INFO

Article history:
Received 8 June 2009
Available online 18 March 2010

AMS subject classification:
62G08
62G09
62G10
62G15
62G20

Keywords:
Local linear modelling
Generalised varying-coefficient models
Hypothesis test
Simultaneous confidence band
Generalised likelihood ratio test

ABSTRACT

Generalised varying-coefficient models (GVC) are very important models. There are a considerable number of literature addressing these models. However, most of the existing literature are devoted to the estimation procedure. In this paper, we systematically investigate the statistical inference for GVC, which includes confidence band as well as hypothesis test. We establish the asymptotic distribution of the maximum discrepancy between the estimated functional coefficient and the true functional coefficient. We compare different approaches for the construction of confidence band and hypothesis test. Finally, the proposed statistical inference methods are used to analyse the data from China about contraceptive use there, which leads to some interesting findings.

1. Introduction

1.1. Preamble

Statistical analysis is always based on some model assumptions. The simplest models are linear models. However, the linearity assumption may not hold in reality. It would suffer from bias if the linear models are used when the linearity assumption does not hold. There are many other parametric models proposed to make the modelling more flexible, however, each have their own limitations. The fully nonparametric modelling makes few model assumptions, but may suffer from variance as it may fail to make use of some prior information available. When the dimension of the covariate is larger, the fully nonparametric modelling would be impracticable due to "curse of dimensionality". A possible way out is to relax the assumptions imposed on the parametric models and make them more flexible, which leads to semiparametric modelling. The commonly used semiparametric models include the additive models [1–5], low-dimensional interaction model [6,7], multiple-index models [8,9], partially linear models [10,11], and their hybrids [12,3].

An important alternative to the additive and other models is the varying-coefficient models [13–18], in which the coefficients of the linear models are replaced by smooth nonparametric functions and hence the regression coefficients are allowed to vary as functions of other factors. The varying-coefficient models are particularly useful in exploring dynamic pattern in many scientific areas.

* Corresponding author.
E-mail address: W.Zhang@bath.ac.uk (W. Zhang).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.03.003

1.2. A motivating example

This paper is stimulated by the data set from China about the contraceptive use there during January 1980 to July 1988. The women were then encouraged to use contraceptive to postpone giving birth. The women's attitude towards contraceptive use in China is varying across different age groups, levels of education, occupations, ethnic groups. Also, whether a woman previously used the contraceptive may affect her failure rate of contraceptive. It is noticeable that the women with previous failure of contraceptive tend to more likely to fail in their contraceptive. Back more than twenty years ago, there were many campaigns organised by the government to advocate late marriage and late birth in China. Some women were self-motivated to use contraceptive, but some were just for response to the campaigns.

There are many factors contributing to the failure rate of contraceptive in China. To explore how the factors affect the failure rate, traditionally the logistic regression models are employed. If we denote the vector of all factors concerned by $\mathbf{X}$, the failure rate of contraceptive by $\pi(\mathbf{X})$, this leads to the following standard logistic regression models

$$\log\frac{\pi(\mathbf{X})}{1-\pi(\mathbf{X})} = \mathbf{X}^{\mathrm{T}}\mathbf{a},$$

each component of $\mathbf{a}$ can be interpreted as the impact of the corresponding factor on the failure rate of contraceptive.

The standard logistic regression models imply a hidden assumption the impacts of all factors concerned are constant. This is apparently implausible for the case in China because China has seen dramatic change since 1979. Some impacts must be varying with time, and the dynamic patterns of the impacts are of both interest and importance as they may reveal how the society is changing with time. A sensible way is to take the time effect into account when analysing the data set. To incorporate the time effect into the modelling, we let the coefficient $\mathbf{a}$ change with time U, which leads to

$$\log\frac{\pi(\mathbf{X},U)}{1-\pi(\mathbf{X},U)} = \mathbf{X}^{\mathrm{T}}\mathbf{a}(U), \tag{1.1}$$

$\pi(\mathbf{X},U)$ is the failure rate of contraceptive. Each component of $\mathbf{a}(U)$ can be interpreted as the dynamic pattern of the impact of the corresponding factor on the failure rate of contraceptive. Models (1.1) is a special case of generalised varying-coefficient models (GVC).

In the study of the data set, some very important questions arise: for a specific factor, does this factor really affect the failure rate of contraceptive? If so, is the impact of this factor varying with time significantly, and to what extent it is varying with time? These questions are very important and interesting. To answer these questions is statistically equivalent to doing hypothesis test and constructing confidence band for the coefficient corresponding to the factor.

There are some existing literature addressing GVC, but very few touching the aspect of confidence band which is a very important part of nonparametric inference. In this paper, we will systematically address the confidence band issue and hypothesis test for GVC. We will investigate different approaches to construct the confidence band and hypothesis test.

To describe the models in a more generic term, we suppose U is a covariate of scalar, $\mathbf{X}=(X_1,\ldots,X_p)$ is a p-dimensional covariate, Y is response variable. We do not confine our discussion in the exponential family, rather we assume the log conditional density function of Y given $(U,\mathbf{X})$ is

$$\ell[g^{-1}\{\mathbf{X}^{\mathrm{T}}\mathbf{a}(U)\}, Y], \tag{1.2}$$

where $\mathbf{a}(U)=(a_1(U),\ldots,a_p(U))^{\mathrm{T}}$ is unknown to be estimated, $g(\cdot)$ is a known link function, $\ell(\cdot,\cdot)$ is known as well. If Y is a discrete random variable, we define its density function as its probability function. (1.2) is the model we are going to address in this paper.

The paper is organised as follows. We begin in Section 2 with a description of estimation procedures. In Section 3 we establish the asymptotic properties of the estimators, which include the asymptotic distribution of the maximum discrepancy between the estimated functional coefficient and the true functional coefficient. We discuss different approaches to construct confidence band and hypothesis test in Section 4. Section 5 is devoted to the simulation study to compare different approaches for confidence band or hypothesis test. Finally, in Section 6, we apply the proposed methods to analyse the data set which stimulates this paper, and explore how the impacts of the factors mentioned before on the failure rate of contraceptive change with time.

2. Estimation procedure

Throughout this paper, $(U_i,\mathbf{X}_i^{\mathrm{T}},Y_i)^{\mathrm{T}}$, $i=1,\ldots,n$, is an i.i.d. sample from $(U,\mathbf{X}^{\mathrm{T}},Y)$, and

$$\mathcal{D}=(U_1,\ldots,U_n,\mathbf{X}_1^{\mathrm{T}},\ldots,\mathbf{X}_n^{\mathrm{T}})^{\mathrm{T}}.$$

For any function/functional vector $g(u)$, we use $g^{(k)}(u)$ to denote its kth derivative. For any k,

$$\mu_k=\int u^kK(u)\mathrm{d}u,\qquad \nu_k=\int u^kK^2(u)\mathrm{d}u.$$

We use $\mathbf{I}_p$ to denote a size p identity matrix, $\mathbf{0}_p$ a size p matrix with each entry being 0, $e_{i,j}$ a vector of length j with the ith component being 1 others being 0, $\otimes$ the Kronecker product, and $f(\cdot)$ the density function of U.

Let $m(u, \mathbf{x})$ be the mean regression function of the response variables Y given the covariates $U = u$ and $\mathbf{X} = \mathbf{x}$, then the generalised varying-coefficient model has the form

$$\eta(u, \mathbf{x}) = g\{m(u, \mathbf{x})\} = \sum_{i=1}^{p} a_j(u)x_j.$$

Define

$$\Gamma(u) = \mathrm{E}\{\rho(U, \mathbf{X})\mathbf{X}\mathbf{X}^T | U = u\} \quad \text{and} \quad q_j(s, y) = (\partial^j/\partial s^j)\ell\{g^{-1}(s), y\},$$

where

$$\rho(u, \mathbf{x}) = -q_2[g\{m(u, \mathbf{x})\}, m(u, \mathbf{x})]. \tag{2.1}$$

Note that $q_k(s, y)$ is linear in y for fixed s, and

$$q_1[g\{m(u, \mathbf{x})\}, m(u, \mathbf{x})] = 0. \tag{2.2}$$

2.1. Local maximum likelihood estimation

For any given u, by the Taylor expansion, we have

$$\mathbf{a}(U_i) \approx \mathbf{a}(u) + h\mathbf{a}^{(1)}(u)\frac{U_i - u}{h},$$

when U_i is in a small neighbourhood of u. This leads to the local log-likelihood function

$$L(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} \ell\left[g^{-1}\left\{\mathbf{X}_i^{\mathrm{T}}\mathbf{a} + \mathbf{X}_i^{\mathrm{T}}\mathbf{b}\frac{U_i - u}{h}\right\}, Y_i\right] K_h(U_i - u), \tag{2.3}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is the kernel function, usually taken to be the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$ due to its minmax property, see [19]. h is the bandwidth.

Let $(\hat{\mathbf{a}}^{\mathrm{T}}, \hat{\mathbf{b}}^{\mathrm{T}})$ maximise $L(\mathbf{a}, \mathbf{b})$. The local maximum likelihood estimator $\hat{\mathbf{a}}(u)$ of $\mathbf{a}(u)$ is taken to be $\hat{\mathbf{a}}$. The estimator $\hat{\mathbf{a}}(u)$ is asymptotically normal and efficient, see [20]. Although the estimator $\hat{\mathbf{a}}(u)$ enjoys many good properties, the implementation of the estimation can be difficult as the computation involved in the maximisation of the local log-likelihood function can be very expensive. To ease the computation burden, Cai et al. [20] proposed an one-step algorithm to compute the estimator. The one-step algorithm dramatically reduces the computation involved in the maximisation, and makes the local maximum likelihood estimation more practicable.

2.2. Estimation for bias

Throughout this paper, expectation, variance or covariance means the conditional expectation, variance or covariance given $\mathcal{D}$.

In this section, we are going to propose an ad hoc estimation procedure for the bias of the estimator $\hat{\mathbf{a}}(u)$. The proposed estimation is based on the asymptotic bias of $\hat{\mathbf{a}}(u)$. Cai et al. [20] have proved the asymptotic bias of $\hat{\mathbf{a}}(u)$ is $2^{-1}h^2\mu_2\mathbf{a}^{(2)}(u)(1 + o_P(1))$. Based on this result, we propose the following estimator of the bias of $\hat{\mathbf{a}}(u)$

$$\widehat{\mathrm{bias}}(\hat{\mathbf{a}}(u)|\mathcal{D}) = 2^{-1}h^2\mu_2\hat{\mathbf{a}}^{(2)}(u). \tag{2.4}$$

The estimator $\hat{\mathbf{a}}^{(2)}(u)$ of $\mathbf{a}^{(2)}(u)$ can be obtained by local cubic maximum likelihood estimation with an appropriate pilot bandwidth ($h_* = O(n^{-1/9})$). The pilot bandwidth h_* can be chosen by the residual squares criterion (RSC) proposed by Fan and Gijbels [19].

More sophistic estimation for the bias could be developed based on the pre-asymptotic substitution idea of Fan and Gijbels [19]. However, in general, it is practically difficult to accurately estimate the bias of $\hat{\mathbf{a}}(u)$ due to poor estimation of higher order derivative of $\mathbf{a}(u)$. The estimation of bias is only of theoretical importance. In the construction of confidence band, an alternative approach to deal with the bias is to use a slightly smaller bandwidth to make the bias ignorable.

2.3. Estimation for variance

The estimation of variance is inevitable when constructing confidence band or hypothesis test. We will appeal the sandwich method to estimate the covariance matrix of $\hat{\mathbf{a}}(u)$. Heuristically, letting $\mathbf{a}$ and $\mathbf{b}$ be $\mathbf{a}(u)$ and $h\mathbf{a}^{(1)}(u)$ respectively in the local log-likelihood function $L(\mathbf{a}, \mathbf{b})$ in (2.3), and applying the Taylor expansion to $L(\mathbf{a}, \mathbf{b})$, we have

$$\begin{aligned} n^{1/2}(\hat{\mathbf{a}}(u) - \mathbf{a}) &\approx -(\mathbf{I}_p, \mathbf{0}_p)\{n^{-1}\ddot{L}(\mathbf{a}, \mathbf{b})\}^{-1}n^{-1/2}\dot{L}(\mathbf{a}, \mathbf{b}) \\ &\approx -(\mathbf{I}_p, \mathbf{0}_p)[\mathrm{E}\{n^{-1}\ddot{L}(\mathbf{a}, \mathbf{b})|\mathcal{D}\}]^{-1}n^{-1/2}\dot{L}(\mathbf{a}, \mathbf{b}), \end{aligned}$$

where $\dot{L}(\mathbf{a}, \mathbf{b})$ and $\ddot{L}(\mathbf{a}, \mathbf{b})$ are respectively the first and second derivative of $L(\mathbf{a}, \mathbf{b})$ with respect to $(\mathbf{a}^{\mathrm{T}}, \mathbf{b}^{\mathrm{T}})$. This leads to

$$\mathrm{cov}(\hat{\mathbf{a}}(u)|\mathcal{D}) \approx (\mathbf{I}_p, \mathbf{0}_p)[\mathrm{E}\{\ddot{L}(\mathbf{a}, \mathbf{b})|\mathcal{D}\}]^{-1}\mathrm{cov}\{\dot{L}(\mathbf{a}, \mathbf{b})|\mathcal{D}\}[\mathrm{E}\{\ddot{L}(\mathbf{a}, \mathbf{b})|\mathcal{D}\}]^{-1}(\mathbf{I}_p, \mathbf{0}_p)^{\mathrm{T}}.$$

Let

$$\Delta_i = \left(1, \frac{U_i - u}{h}\right)^{\mathrm{T}} \left(1, \frac{U_i - u}{h}\right), \qquad q_k(t, y) = (\partial^k/\partial t^k)\ell\{g^{-1}(t), Y\}.$$

By simple calculation, it is easy to see

$$\ddot{L}(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} q_2[\mathbf{X}_i^{\mathrm{T}}(\mathbf{a} + \mathbf{b}(U_i - u)/h), Y_i]\Delta_i \otimes (\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}})K_h(U_i - u)$$

and

$$\mathrm{cov}\{\dot{L}(\mathbf{a}, \mathbf{b})|\mathcal{D}\} = \mathrm{E}\{\Lambda(\mathbf{a},\ \mathbf{b})|\mathcal{D}\}$$

with

$$\Lambda(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} q_1^2[\mathbf{X}_i^{\mathrm{T}}(\mathbf{a} + \mathbf{b}(U_i - u)/h), Y_i]\Delta_i \otimes (\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}})K_h^2(U_i - u).$$

A reasonable estimator of $\mathrm{E}\{\ddot{L}(\mathbf{a}, \mathbf{b})|\mathcal{D}\}$ is $\ddot{L}(\hat{\mathbf{a}}, \hat{\mathbf{b}})$, and a reasonable estimator of $\mathrm{E}\{\Lambda(\mathbf{a}, \mathbf{b})|\mathcal{D}\}$ is $\Lambda(\hat{\mathbf{a}}, \hat{\mathbf{b}})$. So we have the estimator of covariance matrix of $\hat{\mathbf{a}}(u)$

$$\widehat{\mathrm{cov}}\left(\hat{\mathbf{a}}(u)|\mathcal{D}\right) = (\mathbf{I}_p, \mathbf{0}_p)\ddot{L}^{-1}(\hat{\mathbf{a}}, \hat{\mathbf{b}})\Lambda(\hat{\mathbf{a}}, \hat{\mathbf{b}})\ddot{L}^{-1}(\hat{\mathbf{a}}, \hat{\mathbf{b}})(\mathbf{I}_p, \mathbf{0}_p)^{\mathrm{T}}. \tag{2.5}$$

(2.5) is the sandwich estimator of the covariance matrix of $\hat{\mathbf{a}}(u)$.

2.4. Estimation for constant coefficient

When testing whether a specific coefficient in model (1.2) is a constant or not, we need to estimate this coefficient when it is a constant. It would not be sensible to treat this constant as a special case of function and apply the estimation described in Section 2.1 to estimate it, because that would lose the information that the coefficient is a constant, which would eventually pay a price on the variance of the resulting estimator.

Suppose $a_j(\cdot)$ is a constant C_j. Our estimation procedure for C_j is quite simple. It consists of two steps. In the first step, we treat C_j as a function and apply the estimation stated in Section 2.1 to estimate $C_j(U_i)$, $i = 1, \ldots, n$. The resulting estimators are denoted by $\tilde{C}_j(U_i)$, $i = 1, \ldots, n$. In the second step, we average $\tilde{C}_j(U_i)$ over $i = 1, \ldots, n$. The final estimator $\hat{C}_j$ of C_j is taken to be this average

$$\hat{C}_j = n^{-1}\sum_{i=1}^{n}\tilde{C}_j(U_i).$$

In the first step, we treat C_j as a function to estimate, the resulting estimator may have larger variance, however the variance will be reduced by the average in the second step. The final estimator will not suffer on variance side. The idea behind this two-steps estimation is that we use a smaller bandwidth in the first step to control the bias, and the average in the second step to reduce variance. This two-steps estimation for constant coefficient is quite simple but works very well. We will show the estimator obtained by this two-steps estimation is asymptotic normal with convergence rate of $O_P(n^{-1/2})$ when the bandwidth is in a reasonable range, and this estimator is not very sensitive to the bandwidth, it works very well as long as the bandwidth is not too small, nor too large.

When $a_j(\cdot)$ is a constant C_j, the estimation for the functional coefficients $a_i(\cdot)$, $i \neq j$, in model (1.2) is as follows: We first replace the C_j in model (1.2) by its estimator $\hat{C}_j$, then apply the estimation in Section 2.1 to estimate the functional coefficients $a_i(\cdot)$, $i \neq j$. The resulting estimators are denoted by $\tilde{a}_i(\cdot)$, $i \neq j$.

Because the estimator $\hat{C}_j$ is of convergence rate $O_P(n^{-1/2})$, the substitution of $\hat{C}_j$ for C_j in model (1.2) will have little impact on the estimation of $a_i(\cdot)$, $i \neq j$, which implies the estimator $\tilde{a}_i(\cdot)$ will be as good as the estimator of $a_i(\cdot)$ obtained under the condition that C_j is known.

The two-steps estimation idea also appeared in [21,22], though the models there are slightly simpler.

2.5. Bandwidth selection

Bandwidth selection is always an issue in kernel smoothing based nonparametric statistics. The larger bandwidth may gain on variance side but lose on bias side, smaller bandwidth may gain on bias side but lose on variance. An appropriate bandwidth is imperative for a good estimator. There are many criteria for bandwidth selection, see [19]. In this paper, the bandwidth is selected by the following cross validation criterion.

For each i, $i = 1, \ldots, n$, we delete the ith observation $(U_i, \mathbf{X}_i^{\mathrm{T}}, Y_i)$, and estimate $\mathbf{a}(U_i)$ based on the rest observations by the estimation described in Section 2.1 with bandwidth h. The resulting estimator is denoted by $\hat{\mathbf{a}}^{\setminus i}(U_i)$. The log conditional density function of Y at Y_i given $U = U_i$ and $\mathbf{X} = \mathbf{X}_i$ can be estimated by

$$\ell[g^{-1}\{\mathbf{X}_i^{\mathrm{T}}\hat{\mathbf{a}}^{\setminus i}(U_i)\}, Y_i],$$

which naturally leads to the cross validation sum

$$\mathrm{CV}(h) = \sum_{i=1}^{n} \ell[g^{-1}\{\mathbf{X}_i^{\mathrm{T}}\hat{\mathbf{a}}^{\setminus i}(U_i)\}, Y_i].$$

The selected bandwidth is the one maximising $\mathrm{CV}(h)$.

3. Asymptotic properties

In this section, we state our main asymptotic results, and leave the proofs in Appendix. Without loss of generality, we only consider the estimator of the last component $a_p(\cdot)$ of $\mathbf{a}(\cdot)$.

We first present the asymptotic property of the two-steps estimation for the constant coefficient.

Theorem 1. *Under the conditions* (C1)–(C6) *in the Appendix, when $a_p(u)$ is a constant C_p, if $h \to 0$ and $nh^2/(-\log h) \to \infty$, then*

$$\sqrt{n}(\hat{C}_p - C_p) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \mathrm{E}\{e_{p,p}^T \Gamma^{-1}(U) e_{p,p}\}$.

Theorem 1 shows the estimator $\hat{C}_p$ is asymptotic normal with convergence rate of $O_P(n^{-1/2})$ when the bandwidth is in a reasonable wider range.

Let Ψ_q be a $q \times q$ matrix with the (i, j)th element μ_{i+j-2}, $\tilde{\Psi}_q$ be the Ψ_q with μ_i being replaced by ν_i. We also define

$$K_1(t) = e_{1,2}^T \Psi_2^{-1}(1, t)^T K(t), \qquad \nu_{1,0} = \int K_1^2(t)\mathrm{d}t, \qquad r_p(u) = e_{p,p}^T \Gamma^{-1}(u) e_{p,p}.$$

Let $\widehat{\mathrm{bias}}(\hat{a}_j(u)|\mathcal{D})$ be the jth component of $\widehat{\mathrm{bias}}(\hat{\mathbf{a}}(u)|\mathcal{D})$, $\widehat{\mathrm{var}}(\hat{a}_j(u)|\mathcal{D})$ be the jth element on the diagonal of $\widehat{\mathrm{cov}}(\hat{\mathbf{a}}(u)|\mathcal{D})$. We have the following theorem which gives the asymptotic distribution of the maximum discrepancy between the estimated functional coefficient and the true functional coefficient.

Theorem 2. *Under the assumptions* (C1)–(C6) *in the Appendix and $h = O(n^{-b})$, $1/5 \le b < 1 - 2/s$, we have*

$$P\left\{(-2\log h)^{1/2}\left(\nu_{1,0}^{-1/2}\sup_{u\in[0,1]}\left|(nhr_p^{-1}(u)f(u))^{1/2}\left(\hat{a}_p(u) - a_p(u) - \frac{h^2\mu_2}{2}a_p''(u)\right)\right| - d_{v,n}\right) < x\right\}$$
$$\longrightarrow \exp\{-2\exp(-x)\}, \tag{3.1}$$

and

$$P\left\{(-2\log h)^{1/2}\left(\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}\left(\hat{a}_p(u) - a_p(u) - \widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D})\right)\right| - d_{v,n}\right) < x\right\}$$
$$\longrightarrow \exp\{-2\exp(-x)\}, \tag{3.2}$$

where $d_{v,n}$ is the d_n in Lemma A.2 in the Appendix with ν_0 and $K(t)$ being replaced by $\nu_{1,0}$ and $K_1(t)$ respectively.

Remark 1. If the supremum in Theorem 2 is taken on an interval $[c, d]$ instead of $[0, 1]$, Theorem 2 continues to hold under suitable conditions, by using transformation arguments. The results reads as follows:

$$P\Big\{(-2\log\{h/(d-c)\})^{1/2}\Big(\nu_{1,0}^{-1/2}\sup_{u\in[c,d]}|(nhr_p^{-1}(u)f(u))^{1/2}(\hat{a}_p(u) - a_p(u) - \mathrm{bias}(\hat{a}_p(u)|\mathcal{D}))| - \tilde{d}_{v,n}\Big) < x\Big\}$$
$$\longrightarrow \exp\{-2\exp(-x)\},$$

where $\tilde{d}_{v,n}$ is the $d_{v,n}$ in Theorem 2 with h being replaced by $h/(d-c)$.

Theorem 3. *If $nh^2/(-\log h) \to \infty$, under the conditions of Theorem 2, when $a_p(\cdot)$ is a constant C_p, we have*

$$P\left\{(-2\log h)^{1/2}\left(\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(a_p(u)|\mathcal{D})\}^{\frac{1}{2}}}\left(\hat{a}_p(u) - \hat{C}_p - \widehat{\mathrm{bias}}(a_p(u)|\mathcal{D})\right)\right| - d_{v,n}\right) < x\right\} \longrightarrow \exp\{-2\exp(-x)\}.$$

Theorem 3 can be directly used to test the hypothesis that $a_p(\cdot)$ is a constant.

4. Confidence band and hypothesis test

In this section, we will investigate a few approaches to construct confidence band and hypothesis test. We will address confidence band first, then hypothesis test. Without loss of generality, we will focus our discussion on the last component $a_p(\cdot)$ of $\mathbf{a}(\cdot)$ in model (1.2).

4.1. Confidence band

The construction of confidence band is based on the distribution of the maximum discrepancy between the estimated functional coefficient and the true functional coefficient. It is hard to find the exact distribution of the maximum discrepancy, however, it can be estimated by either its asymptotic form or bootstrap. We will discuss these two approaches respectively. Without loss of generality, we will focus on the construction of the confidence band on the interval [0, 1].

4.1.1. Asymptotic distribution based approach

The construction of confidence band based on the asymptotic distribution is quite straightforward. The Theorem 2 in Section 3 gives the following $1-\alpha$ confidence band of $a_p(\cdot)$ on the interval $[0,\ 1]$:

$$\hat{a}_p(u) - \widehat{\text{bias}}(\hat{a}_p(u)|\mathcal{D}) \pm \Delta_\alpha(u),$$

where $\widehat{\text{bias}}(\hat{a}_p(u)|\mathcal{D})$ is the pth component of $\widehat{\text{bias}}\left(\hat{\mathbf{a}}(u)|\mathcal{D}\right)$ in (2.4), and

$$\Delta_\alpha(u) = \left(d_{v,n} + [\log 2 - \log\{-\log(1-\alpha)\}]\,(-2\log h)^{-1/2}\right)\left\{\widehat{\text{var}}(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2},$$

with $\widehat{\text{var}}(\hat{a}_p(u)|\mathcal{D})$ being the pth element on the diagonal of $\widehat{\text{cov}}(\hat{\mathbf{a}}(u)|\mathcal{D})$ in (2.5).

The interpretation of this confidence band is that the probability of the true curve $a_p(u)$ sandwiched between the curves $\hat{a}_p(u) - \widehat{\text{bias}}(\hat{a}_p(u)|\mathcal{D}) - \Delta_\alpha(u)$ and $\hat{a}_p(u) - \widehat{\text{bias}}(\hat{a}_p(u)|\mathcal{D}) + \Delta_\alpha(u)$ on the interval $[0,\ 1]$ is $1-\alpha$.

The advantage of the asymptotic distribution based approach is it is easy to implement and the computation involved is very cheap. However, when sample size is moderate, the coverage probability of the resulting confidence band may not be as good as expected.

4.1.2. Bootstrap based approach

The bootstrap is a very useful tool for statistical inference. We are now going to describe how to use it to construct confidence band for $a_p(\cdot)$. Let

$$T = \sup_{u\in[0,1]} \frac{\left|\hat{a}_p(u) - a_p(u)\right|}{\left\{\text{var}(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2}}.$$

Suppose the upper α quantile of T is c_α. If both c_α and $\text{var}(\hat{a}_p(u)|\mathcal{D})$ were known, the confidence band of $a_p(\cdot)$ on the interval [0, 1] would be

$$\hat{a}_p(u) \pm \left\{\text{var}(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2} c_\alpha. \tag{4.1}$$

However c_α and $\text{var}(\hat{a}_p(u)|\mathcal{D})$ unknown. We will estimate them by bootstrap later. Suppose we have the estimators $\hat{c}_\alpha$ and $\text{var}^*(\hat{a}_p(u)|\mathcal{D})$ of c_α and $\text{var}(\hat{a}_p(u)|\mathcal{D})$. Substituting $\hat{c}_\alpha$ and $\text{var}^*(\hat{a}_p(u)|\mathcal{D})$ for c_α and $\text{var}(\hat{a}_p(u)|\mathcal{D})$ respectively in (4.1) leads to the $1-\alpha$ confidence band of $a_p(\cdot)$

$$\hat{a}_p(u) \pm \left\{\text{var}^*(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2} \hat{c}_\alpha.$$

We are now turning to demonstrate how to estimate c_α and $\text{var}(\hat{a}_p(u)|\mathcal{D})$ by bootstrap. The whole estimation procedure consists of the following five steps.

(1) Estimate $\mathbf{a}(\cdot)$ by the estimation method in Section 2.1. Denote the resulting estimator by $\hat{\mathbf{a}}(\cdot)$.
(2) For each i, $i = 1, \ldots, n$, generate a bootstrap sample member Y_i^* based on the log conditional density function

$$\ell\left[g^{-1}\left\{\mathbf{X}_i^{\mathrm{T}}\hat{\mathbf{a}}(U_i)\right\}, Y_i\right].$$

Estimate $\mathbf{a}(\cdot)$ by the estimation method proposed in Section 2.1 based on the bootstrap sample $(U_i, \mathbf{X}_i^{\mathrm{T}},\ Y_i^*)$, $i = 1, \ldots, n$. Denote the resulting estimator by $\hat{\mathbf{a}}^*(\cdot)$ and term it a bootstrap sample member of $\hat{\mathbf{a}}(\cdot)$.
(3) Repeat (2) m times to get a size m bootstrap sample of $\hat{\mathbf{a}}(\cdot)$: $\hat{\mathbf{a}}_i^*(\cdot)$, $i = 1, \ldots, m$. The estimator $\text{cov}^*(\hat{\mathbf{a}}(\cdot))$ of $\text{cov}(\hat{\mathbf{a}}(\cdot))$ is taken to be the sample covariance of $\hat{\mathbf{a}}_i^*(\cdot)$, $i = 1, \ldots, m$, and the pth element on the diagonal of $\text{cov}^*(\hat{\mathbf{a}}(\cdot))$ is the estimator $\text{var}^*(\hat{a}_p(\cdot)|\mathcal{D})$.

(4) Repeat (2) M times to get a bootstrap sample of size M for $\hat{\mathbf{a}}(\cdot)$: $\hat{\mathbf{a}}_i^*(\cdot)$, $i = 1, \ldots, M$. Compute

$$T_i^* = \sup_{u \in \mathcal{D}} \frac{\left|\hat{a}_{i,p}^*(u) - \hat{a}_p(u)\right|}{\left\{\mathrm{var}^*(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2}}, \quad i = 1, \ldots, M,$$

where $\hat{a}_{i,p}^*(\cdot)$ is the pth component of $\hat{\mathbf{a}}_i^*(\cdot)$. We term T_i^*, $i = 1, \ldots, M$, the bootstrap sample of T.

(5) Use the upper α percentile of T_i^*, $i = 1, \ldots, M$, to estimate the upper α quantile c_α of T.

The coverage probability of the confidence band constructed by the bootstrap approach would be better than that generated by the asymptotic distribution based approach when the sample size is moderate. However, the computation involved in the bootstrap approach is much more expensive than that involved in the asymptotic distribution based approach.

4.2. Hypothesis test

The hypothesis test is another important part of statistical inference. The hypothesis

$$H_0 : a_p(\cdot) = C_p \longleftrightarrow H_1 : a_p(\cdot) \neq C_p \tag{4.2}$$

is also of practical interest. The null hypothesis means the impact of X_p is not varying with time, and the alternative hypothesis means the impact is varying with time. We will in this section discuss three approaches to construct the hypothesis test for the hypothesis (4.2). Without loss of generality, we assume the support set of $a_p(\cdot)$ is $[0, 1]$.

4.2.1. Asymptotic distribution based approach

Under the null hypothesis of (4.2), $a_p(\cdot)$ is a constant C_p. Applying the estimation for constant coefficient in Section 2.4, we obtain the estimator $\hat{C}_p$ of C_p. A natural test statistic is

$$\mathcal{T} = \sup_{u \in [0,1]} \frac{\left|\hat{a}_p(u) - \hat{C}_p - \widehat{\mathrm{bias}}\left(\hat{a}_p(u)|\mathcal{D}\right)\right|}{\left\{\widehat{\mathrm{var}}\left(\hat{a}_p(u)|\mathcal{D}\right)\right\}^{1/2}}.$$

The Theorem 3 in Section 3 leads to the hypothesis test of size α: rejecting the null hypothesis when

$$\mathcal{T} > d_{v,n} + [\log 2 - \log\{-\log(1-\alpha)\}]\,(-2\log h)^{-1/2},$$

accepting the null hypothesis otherwise.

4.2.2. Bootstrap based approach

In this section, we are going to use bootstrap together with the quantity

$$\mathbb{T} = \sup_{u \in [0,1]} \frac{\left|\hat{a}_p(u) - C_p\right|}{\left\{\mathrm{var}(\hat{a}_p(u))|\mathcal{D}\right\}^{1/2}}$$

to construct the hypothesis test for the hypothesis (4.2).

Suppose the upper α quantile of $\mathbb{T}$ under null hypothesis of (4.2) is c_α. If c_α and the C_p and $\mathrm{var}(\hat{a}_p(u)|\mathcal{D})$ in $\mathbb{T}$ were known, we would have the hypothesis test with the rejection region

$$\sup_{u \in [0,1]} \frac{\left|\hat{a}_p(u) - C_p\right|}{\left\{\mathrm{var}(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2}} > c_\alpha. \tag{4.3}$$

Unfortunately, c_α, C_p and $\mathrm{var}(\hat{a}_p(u)|\mathcal{D})$ are unknown.

We first assume that we have the estimators $\hat{c}_\alpha$, $\hat{C}_p$ and $\mathrm{var}^*(\hat{a}_p(u)|\mathcal{D})$ of c_α, C_p and $\mathrm{var}(\hat{a}_p(u)|\mathcal{D})$. We will describe how to get these estimators by bootstrap later. Substituting $\hat{c}_\alpha$, $\hat{C}_p$ and $\mathrm{var}^*(\hat{a}_p(u)|\mathcal{D})$ for c_α, C_p and $\mathrm{var}(\hat{a}_p(u)|\mathcal{D})$ in (4.3), we have the hypothesis test of size α for (4.2): rejecting the null hypothesis when

$$\sup_{u \in [0,1]} \frac{\left|\hat{a}_p(u) - \hat{C}_p\right|}{\left\{\mathrm{var}^*(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2}} > \hat{c}_\alpha,$$

accepting the null hypothesis otherwise.

The $\hat{C}_p$ can be obtained by the estimation in Section 2.4. We are now demonstrating how to get the estimators $\hat{c}_\alpha$ and $\mathrm{var}^*(\hat{a}_p(u)|\mathcal{D})$ by bootstrap. The estimation based on the bootstrap resampling under the null hypothesis of (4.2) consists of the following five steps:

(1) Under the null hypothesis, namely $a_p(\cdot) = C_p$, we estimate C_p and the functional coefficients $a_j(\cdot), j = 1, \ldots, p-1$, by the estimation in Section 2.4. The resulting estimators are denoted by $\hat{C}_p$ and $\tilde{a}_j(\cdot), j = 1, \ldots, p-1$, respectively.

(2) Let X_{ij} be the jth component of $\mathbf{X}_i$. For each i, $i = 1, \ldots, n$, generate a bootstrap sample member Y_i^* based on the log conditional density function

$$\ell\left[g^{-1}\left\{\sum_{j=1}^{p-1} X_{ij}\tilde{a}_j(U_i) + X_{ip}\hat{C}_p\right\}, Y\right].$$

Treat $a_p(\cdot)$ as a function and estimate it by the estimation in Section 2.1 based on the bootstrap sample $(U_i, \mathbf{X}_i^{\mathrm{T}}, Y_i^*)$, $i = 1, \ldots, n$. Denote the resulting estimator by $\hat{a}_p^*(\cdot)$, and term it a bootstrap sample member of $\hat{a}_p(\cdot)$.

(3) Repeat (2) m times to get a size m bootstrap sample, $\hat{a}_{i,p}^*(\cdot)$, $i = 1, \ldots, m$, of $\hat{a}_p(\cdot)$. The estimator $\mathrm{var}^*(\hat{a}_p(\cdot)|\mathcal{D})$ of $\mathrm{var}(\hat{a}_p(\cdot)|\mathcal{D})$ is taken to be the sample variance of $\hat{a}_{i,p}^*(\cdot)$, $i = 1, \ldots, m$.

(4) Repeat (2) M times to get a size M bootstrap sample, $\hat{a}_{i,p}^*(\cdot)$, $i = 1, \ldots, M$, of $\hat{a}_p(\cdot)$. Compute

$$\mathbb{T}_i^* = \sup_{u\in[0,1]} \frac{\left|\hat{a}_{i,p}^*(u) - \hat{C}_p\right|}{\left\{\mathrm{var}^*(\hat{a}_p(u)|\mathcal{D})\right\}^{1/2}}, \quad i = 1, \ldots, M.$$

We term $\mathbb{T}_i^*$, $i = 1, \ldots, M$, a size M bootstrap sample of $\mathbb{T}$.

(5) The estimator $\hat{c}_\alpha$ of c_α is taken to be the upper α percentile of $\mathbb{T}_i^*$, $i = 1, \ldots, M$.

4.2.3. Generalised maximum likelihood ratio approach

The generalised maximum likelihood ratio test (GMLRT) proposed by Fan et al. [23] is a powerful tool for nonparametric hypothesis test. It has been widely used in many statistical models, see [24]. The generalised maximum likelihood ratio test statistic for our setting can be constructed as

$$\mathcal{R} = \sum_{i=1}^{n} \ell\left[g^{-1}\left\{\mathbf{X}_i^{\mathrm{T}}\hat{\mathbf{a}}(U_i)\right\}, Y_i\right] - \sum_{i=1}^{n} \ell\left[g^{-1}\left\{\sum_{j=1}^{p-1} X_{ij}\tilde{a}_j(U_i) + X_{ip}\hat{C}_p\right\}, Y_i\right],$$

$\hat{\mathbf{a}}(\cdot)$ is the estimator of $\mathbf{a}(\cdot)$ obtained by the estimation in Section 2.1. $\hat{C}_p$ and $\tilde{a}_j(\cdot), j = 1, \ldots, p-1$, are the estimators of C_p and $a_j(\cdot)$ obtained by the estimation in Section 2.4 when the null hypothesis of (4.2) holds. The statistic $\mathcal{R}$ can be roughly viewed as the difference on the maximum of the log likelihood function between without and with assuming the null hypothesis of (4.2) holds. It is clear that the larger the $\mathcal{R}$, the more unlikely the null hypothesis of (4.2) holds. Let c_α be the upper α quantile of $\mathcal{R}$ under the null hypothesis of (4.2), the size α GMLRT for the hypothesis (4.2) is: we reject the null hypothesis when $\mathcal{R} > c_\alpha$, accept the null hypothesis otherwise.

It is almost impossible to find the exact distribution of $\mathcal{R}$. To find c_α, there are two ways, one is to count on the asymptotic distribution of $\mathcal{R}$, another is bootstrap approach. We are now presenting the asymptotic distribution of $\mathcal{R}$.

Theorem 4. *Suppose the conditions* (C1)–(C6) *in the Appendix hold. Under the null hypothesis of* (4.2), *when* $nh^6 \to 0$, $nh^2/(\log n)^2 \to \infty$,

$$\sigma_n^{-1}(\mathcal{R} - \mu_n) \xrightarrow{D} \mathcal{N}(0, 1)$$

where

$$\mu_n = \frac{1}{h}\left\{K(0) - \frac{1}{2}\int K^2(t)\mathrm{d}t\right\}, \qquad \sigma_n^2 = \frac{2}{h}\int\left\{K(t) - \frac{1}{2}K * K(t)\right\}^2 \mathrm{d}t$$

$K * K$ is the convolution of K.

Remark 2. Theorem 4 implies that under the null hypothesis of (4.2) and the conditions (C1)–(C6) in the Appendix, $r_K\mathcal{R} \xrightarrow{D} \chi^2_{r_K\mu_n}$, where

$$r_K = \frac{K(0) - \frac{1}{2}\int K^2(t)\mathrm{d}t}{\int\left\{K(t) - \frac{1}{2}K * K(t)\right\}^2 \mathrm{d}t}.$$

Theorem 4 suggests that the asymptotic distribution of $\mathcal{R}$ is free of the unknown parameter and nuisance functions under the null hypothesis. This is the so called Wilks phenomenon, which is quite important. It proves that the $\mathcal{R}$ is indeed a test statistic, and the c_α can be approximated by $\sigma_n z_\alpha + \mu_n$. z_α is the upper α quantile of the standard normal distribution.

The c_α can also be estimated by the bootstrap approach. Usually, to generate the bootstrap sample under null hypothesis we need to know the unknown parameters involved in the model when the null hypothesis holds. However, for our case, it

Table 1
Coverage probabilities based on 1000 simulations when sample size is 1000.

$1-\alpha$	$a_1(\cdot)$		$a_2(\cdot)$	
	Method one	Method two	Method one	Method two
0.99	0.985	0.992	0.993	0.994
0.95	0.914	0.961	0.898	0.949
0.90	0.809	0.900	0.779	0.920

Method one is the method based on asymptotic distribution, method two is the method based on bootstrap.

is no need. This is because the asymptotic distribution of $\mathcal{R}$ under the null hypothesis is free of any unknown parameters. We can just simply assign some reasonable values to the unknown parameters. We recommend to replace the unknown parameters and functions involved in the model by their estimators obtained under null hypothesis when generating the bootstrap sample.

The way to generate the bootstrap sample $(U_i, \mathbf{X}_i^{\mathrm{T}}, Y_i^*)$, $i = 1, \ldots, n$, is exactly the same as that in Section 4.2.2. After the bootstrap sample $(U_i, \mathbf{X}_i^{\mathrm{T}}, Y_i^*)$, $i = 1, \ldots, n$, being generated, based on the generated bootstrap sample we apply the estimation in Section 2.1 to estimate $\mathbf{a}(\cdot)$ without assuming the null hypothesis holds, and the estimation in Section 2.4 to estimate C_p and $a_j(\cdot)$, $j = 1, \ldots, p-1$, under the assumption that the null hypothesis holds. Denote the resulting estimators by $\hat{\mathbf{a}}^*(\cdot)$, $\hat{C}_p^*$ and $\tilde{a}_j^*(\cdot)$, $j = 1, \ldots, p-1$, respectively. A bootstrap sample member $\mathcal{R}^*$ of $\mathcal{R}$ is the $\mathcal{R}$ with $\hat{\mathbf{a}}(\cdot)$, $\hat{C}_p$ and $\tilde{a}_j(\cdot)$, $j = 1, \ldots, p-1$, being replaced by $\hat{\mathbf{a}}^*(\cdot)$, $\hat{C}_p^*$ and $\tilde{a}_j^*(\cdot)$ respectively. Repeating the procedure M times, we get a size M bootstrap sample, $\mathcal{R}_i^*$, $i = 1, \ldots, M$, of $\mathcal{R}$. We use the upper α percentile of $\mathcal{R}_i^*$, $i = 1, \ldots, M$, to estimate the c_α.

5. Simulation study

In this section, we are going to use a simulated example to demonstrate how well the approaches described in Section 4 for confidence band and hypothesis test work. We will also compare different approaches to see which one works best.

Example 1. We generate data $(Y_i, \mathbf{X}_i, U_i)$, $i = 1, \ldots, n$, from the following logistic regression model

$$\log\left\{\frac{P(Y=1|\mathbf{X},U)}{1-P(Y=1|\mathbf{X},U)}\right\} = \mathbf{X}^{\mathrm{T}}\mathbf{a}(U). \tag{5.1}$$

$\mathbf{a}(\cdot) = (a_1(\cdot), a_2(\cdot))^{\mathrm{T}}$. The sample size n is set to be 1000. X_i, $i = 1, \ldots, n$, are independently generated from normal distribution $N(\mathbf{0}, \mathbf{I}_2)$. U_i, $i = 1, \ldots, n$, are independently generated from uniform distribution $U(0, 1)$. We set

$$a_1(u) = \sin(2\pi u), \qquad a_2(u) = \cos(2\pi u).$$

The kernel function $K(t)$ involved in the estimation is taken to be the Epanechnikov kernel $0.75(1-t^2)_+$. The bandwidth is chosen to be 0.15, which is 80% of the average of the bandwidths selected by the proposed CV in Section 2.5 across 100 simulations. The reason for us not to use the bandwidth selected by CV directly is undersmoothing is needed to make the bias negligible.

As the computation involved in computing the CV is very expensive, it becomes unduly time consuming to compute the bandwidth selected by the CV for each simulation when the number of the simulations is very large. That is why we set the bandwidth to be 0.15, which is obtained through 100 simulations, for all simulations. The simulation results show this bandwidth does work well.

The selection of the optimal undersmoothing bandwidth is challenging. Undersmoothing appears quite often in semiparametric modelling. People always use some ad hoc approaches to select bandwidth when undersmoothing is required. Our method to select bandwidth here does not make any difference. Roughly speaking, our selected bandwidth is 80% of the bandwidth selected by CV. It is data-driven and works well in our simulation studies, however, we do not claim it is the best approach. To systemically investigate the optimal undersmoothing bandwidth selection is of great interest and importance, however, it does go beyond the scope of this paper.

5.1. Confidence band

The asymptotic distribution based approach and the bootstrap approach are respectively used to construct the confidence bands of $a_1(\cdot)$ and $a_2(\cdot)$. We conduct 1000 simulations to compute the coverage probability of the confidence band constructed by either the asymptotic distribution based approach or the bootstrap approach when the confidence level $1-\alpha$ is taken to be 90%, 95% and 99% respectively. The results are presented in Table 1. From Table 1, it is easy to see the bootstrap approach works better than the asymptotic distribution based approach. By simple calculation, we have the Monte Carol error of size $\sqrt{0.9 \times 0.1/1000} \approx 0.0095$ for $\alpha = 0.10$, 0.0069 for $\alpha = 0.05$, and 0.0031 for $\alpha = 0.01$. Taking these Monte Carol errors into account, we can see the bootstrap approach is doing quite well.

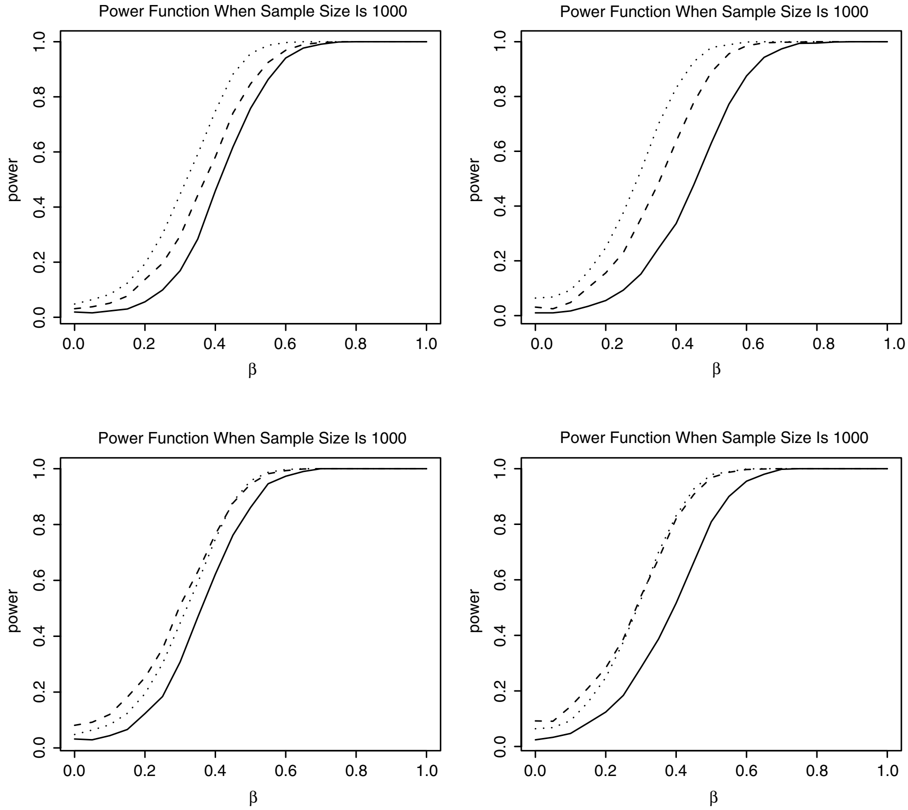

Fig. 1. The upper panel is for significant level $\alpha = 0.05$, and the lower panel is for $\alpha = 0.1$. The left one is for $a_1(\cdot)$, the right one is for $a_2(\cdot)$. The solid line is the power function of the bootstrap based test, the dashed line is the power function of the asymptotic distribution based test, and the dotted line is the power function of the generalised maximum likelihood ratio test.

5.2. Hypothesis test

To examine how powerful the hypothesis tests discussed in Section 4 are, we set the $a_1(u)$ in Example 1 to be

$$a_1(u) = 1 - \beta + \beta \sin(2\pi u).$$

For each fixed β, we conduct 1000 simulations, and in each of these simulations, we generate a sample of size $n = 1000$, then set the significant level $\alpha = 0.05$ and test the hypothesis

$$H_0 : a_1(\cdot) = C_1 \longleftrightarrow H_1 : a_1(\cdot) \neq C_1,$$

and evaluate the power of the test at this fixed β based on the 1000 simulations. When $\beta = 0$, the power becomes the size of the test. The plots of the power function against β for the three hypothesis tests discussed in Section 4.2 are depicted in the upper panel of Fig. 1. The plots suggest the asymptotic distribution based approach performs better than the bootstrap approach, and the GMLRT with the upper α quantile c_α being estimated by bootstrap does best. We can also see that the GMLRT is quite powerful.

To compare the three hypothesis tests more deeply, we set the significant level $\alpha = 0.1$, and compute their power functions. The obtained power functions are presented in the lower panel of Fig. 1, from which we can see that there is little difference between the GMLRT and the asymptotic distribution based test, and both of them are more powerful than the bootstrap one. So, we can safely claim the GMLRT is the most powerful one among the three tests discussed.

We have also done the same exercise to the coefficient $a_2(\cdot)$, and the results are also depicted in Fig. 1. From Fig. 1, we can see the conclusion is exactly the same as that from $a_1(\cdot)$.

We have also set the sample size to be 500 and repeated all we did above. The coverage probabilities of the confidence bands are reported in Table 2, and the power functions of the three hypothesis tests are presented in Fig. 2. Table 2 and Fig. 2 tell us the same story as before, that is bootstrap performs best in the construction of confidence band and the GMLRT is the most powerful one among the three tests we investigate.

To more vigorously compare the proposed two approaches for construction of confidence band and the three hypothesis tests, and examine the effect of different specifications of functional coefficients on the construction of confidence band and hypothesis test, we study another simulated example.

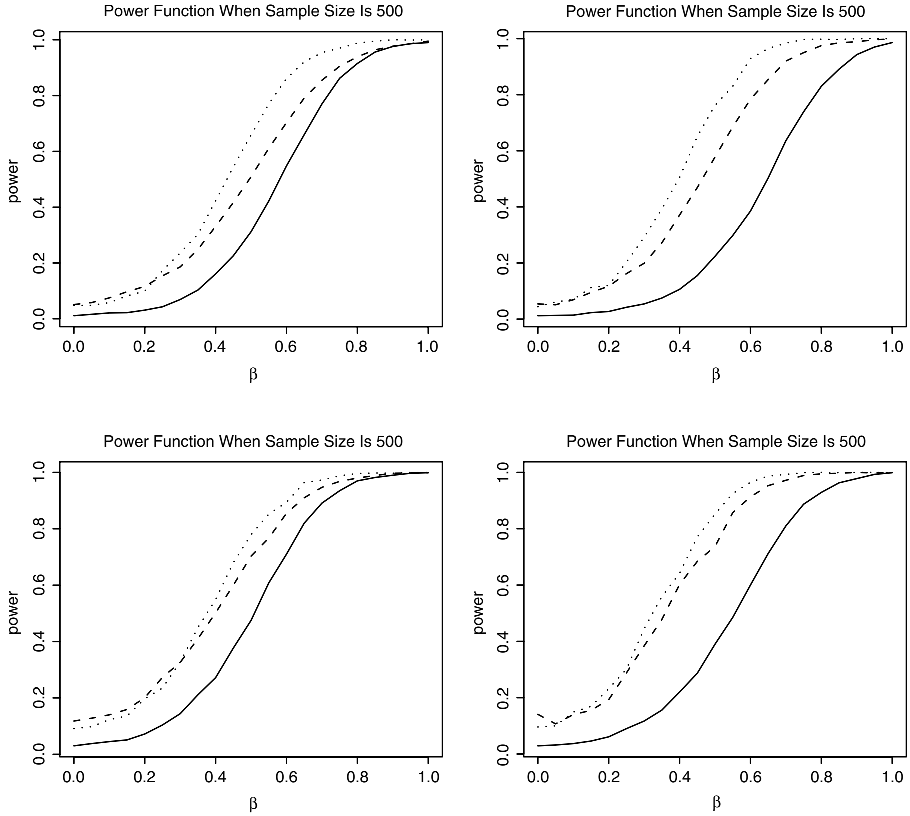

Fig. 2. The upper panel is for significant level $\alpha = 0.05$, and the lower panel is for $\alpha = 0.1$. The left one is for $a_1(\cdot)$, the right one is for $a_2(\cdot)$. The solid line is the power function of the bootstrap based test, the dashed line is the power function of the asymptotic distribution based test, and the dotted line is the power function of the generalised maximum likelihood ratio test.

Table 2
Coverage probabilities based on 1000 simulations when sample size is 500.

$1-\alpha$	$a_1(\cdot)$		$a_2(\cdot)$	
	Method one	Method two	Method one	Method two
0.99	0.987	0.996	0.982	0.997
0.95	0.885	0.951	0.872	0.958
0.90	0.768	0.907	0.785	0.914

Method one is the method based on asymptotic distribution, method two is the method based on bootstrap.

Example 2. To keep consistent with the real data analysis in Section 6, we still generate data $(Y_i, \mathbf{X}_i, U_i)$, $i = 1, \dots, n$, from the logistic regression model (5.1), however, the functional coefficients are set to be

$$a_1(u) = 4u(1-u), \qquad a_2(u) = 0.5(\cos(\pi u) + \sin(\pi u)).$$

$X_i, i = 1, \dots, n$, are independently generated from normal distribution $N(\mathbf{0}, \mathbf{I}_2)$. $U_i, i = 1, \dots, n$, are independently generated from uniform distribution $U(0, 1)$.

Setting sample size to be 1000 or 500, we study the confidence bands of the functional coefficients when the confidence level $1-\alpha$ is taken to be 90%, 95% and 99% respectively, and the power functions of the three hypothesis tests when the significant level is taken to be 0.05 and 0.1 respectively.

To avoid unnecessary replication, we only present the obtained results without any detail. The coverage probabilities of the confidence bands in 1000 simulations are reported in Table 3 when sample size is 1000, and Table 4 when sample size is 500. The power functions of the three tests are presented in Fig. 3 when sample size is 1000, and Fig. 4 when sample size is 500. From Tables 3 and 4, we can see that the bootstrap approach works best in the construction of confidence band which is in line with what we have seen in Example 1. Figs. 3 and 4 show that the GMLRT is the most powerful one among the three tests we study, which is again in line with what we have seen in Example 1.

To summarise the findings from our simulation studies, we conclude that when it comes to construction of confidence band, go for bootstrap; when it comes to hypothesis test, go for the GMLRT.

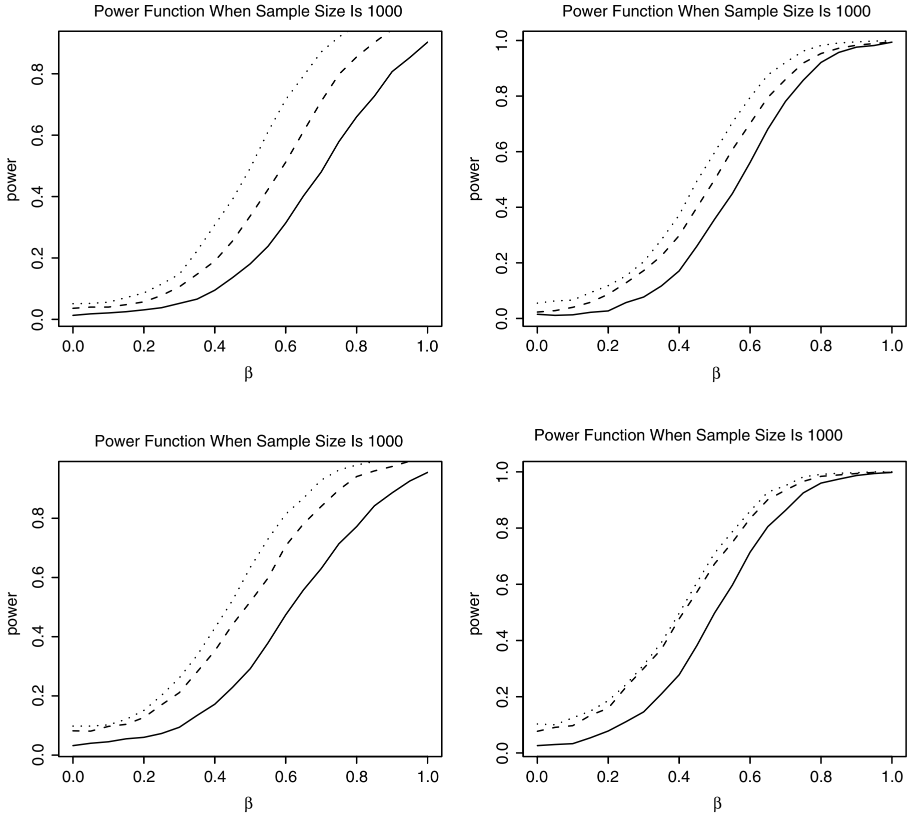

Fig. 3. The upper panel is for significant level $\alpha = 0.05$, and the lower panel is for $\alpha = 0.1$. The left one is for $a_1(\cdot)$, the right one is for $a_2(\cdot)$. The solid line is the power function of the bootstrap based test, the dashed line is the power function of the asymptotic distribution based test, and the dotted line is the power function of the generalised maximum likelihood ratio test.

Table 3
Coverage probabilities based on 1000 simulations when sample size is 1000.

$1-\alpha$	$a_1(\cdot)$		$a_2(\cdot)$	
	Method one	Method two	Method one	Method two
0.99	0.996	0.994	0.999	0.986
0.95	0.952	0.947	0.963	0.955
0.90	0.885	0.901	0.884	0.909

Method one is the method based on asymptotic distribution, method two is the method based on bootstrap.

Table 4
Coverage probabilities based on 1000 simulations when sample size is 500.

$1-\alpha$	$a_1(\cdot)$		$a_2(\cdot)$	
	Method one	Method two	Method one	Method two
0.99	0.993	0.982	0.989	0.986
0.95	0.937	0.964	0.925	0.963
0.90	0.859	0.887	0.828	0.911

Method one is the method based on asymptotic distribution, method two is the method based on bootstrap.

6. Real data analysis

In this section, we are going to use the proposed methods to analyse the data set from China about the contraceptive use there. Women from different backgrounds may have different attitudes towards the contraceptive use. There are many factors affecting the contraceptive use in China, of which we are particularly interested in how the following factors affect the contraceptive use in China: the women's age, type of region of residence, education, occupation, ethnic, previous use of contraceptive, previous failure of contraceptive, and the motivation to contraceptive use.

The women's age is grouped to "less than 24", "25 to 29" (x_2), "30 to 34" (x_3) and "over 35" (x_4). We take "less than 24" as reference, and the differences in the impacts on the contraceptive use among different age groups are modelled by the

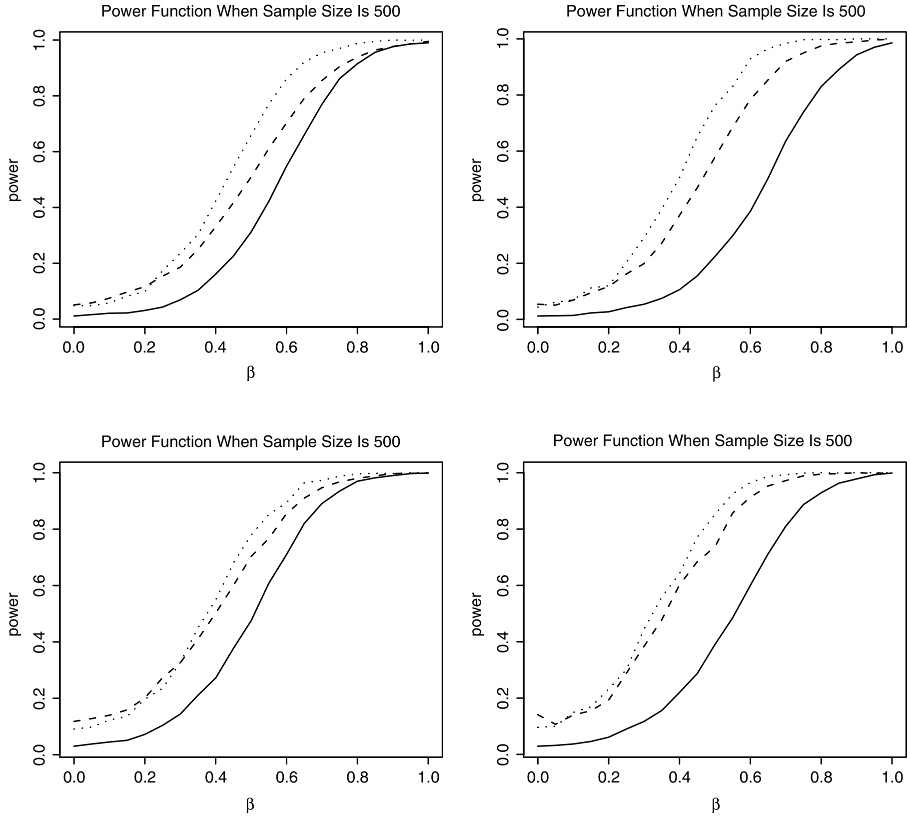

Fig. 4. The upper panel is for significant level $\alpha = 0.05$, and the lower panel is for $\alpha = 0.1$. The left one is for $a_1(\cdot)$, the right one is for $a_2(\cdot)$. The solid line is the power function of the bootstrap based test, the dashed line is the power function of the asymptotic distribution based test, and the dotted line is the power function of the generalised maximum likelihood ratio test.

dummy variables x_i, $i = 2, 3, 4$. We take "urban" as reference, the differences between urban and rural is modelled a dummy variable x_5. The woman's education is categorised as "primary −" or "junior +". We take "primary −" as reference, and the difference between "primary −" or "junior +" is modelled by a dummy variable x_6. We categorise the woman's occupation as "agriculture", "industry" (x_7), "service" (x_8), "professional" (x_9) or "other non-agriculture" (x_{10}). We take "agriculture" as reference, and the differences among different occupations are modelled by the dummy variables x_i, $i = 7, \ldots, 10$. We take "non-Han" as reference, the difference between "Han" and "non-Han" is modelled by a dummy variable x_{11}. We use dummy variables x_{12} and x_{13} to model "previous use of contraceptive" and "previous failure of contraceptive" respectively. The motivation to contraceptive use is categorised as "self motivated" or "response to campaign". We take "self motivated" as reference, and use x_{14} to model the difference between "self motivated" and "response to campaign". Chronological time is denoted by U. We set $x_1 = 1$ to incorporate the intercept into the modelling. The dependent variable, Y, is taken to be 1 if the contraceptive fails, 0 otherwise.

The model (1.1) is used to fit the data set. The kernel function involved in the estimation is still taken to be Epanechnikov kernel, and the bandwidth is chosen to be 11% of the range of U.

As the GMLRT performs best among the three methods for hypothesis test, the GMLRT is chosen to serve the analysis of the data set.

We first apply the GMLRT to test whether the impacts of the factors concerned are changing with time. That is to test the following hypotheses

$$H_j : a_j(\cdot) = C_j \longleftrightarrow H_{j,1} : a_j(\cdot) \neq C_j, \quad j = 1, \ldots, 14. \tag{6.1}$$

The obtained P-values are presented in Table 5. Table 5 suggests $a_1(\cdot)$ and $a_{14}(\cdot)$ are not constant, and $a_j(\cdot), j = 2, \ldots, 13$, can be treated as constants. That means there is a time varying trend in the failure rate of contraceptive, and the impact of the motivation is also varying with time.

We apply the estimation in Section 2.4 to estimate the constant coefficients, and present the results in Table 6. The standard errors of the estimators are obtained by bootstrap and presented in Table 6 too. From Table 6, we can see the coefficient a_i of x_i is not significantly different to zero when $i = 5, \ldots, 12$. This means there is no significant difference on the failure rate of contraceptive between the women in rural area and those in urban area. This also suggests that neither level of education, nor type of occupation, nor ethnic contributes significantly to the failure rate of contraceptive.

Table 5
The P-values for the hypotheses (6.1).

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	H_{11}	H_{12}	H_{13}	H_{14}
P-value	0.000	0.690	0.056	0.272	0.320	0.704	0.235	0.931	0.606	0.833	0.341	0.125	0.965	0.049

$H_j, j = 1, \ldots, 14$, are the null hypotheses in (6.1) in Section 6.

Table 6
The estimates of the constant coefficients.

	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	$\hat{a}_7$	$\hat{a}_8$	$\hat{a}_9$	$\hat{a}_{10}$	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{13}$
Value	−0.341	−1.118	−1.945	0.019	0.038	0.070	−0.269	−0.138	0.067	0.042	−0.059	0.757
SE	0.085	0.165	0.343	0.136	0.086	0.152	0.269	0.259	0.206	0.143	0.129	0.132

The row beginning with value is the estimates of the constant coefficients, and the row beginning with SE is the standard errors of the estimators.

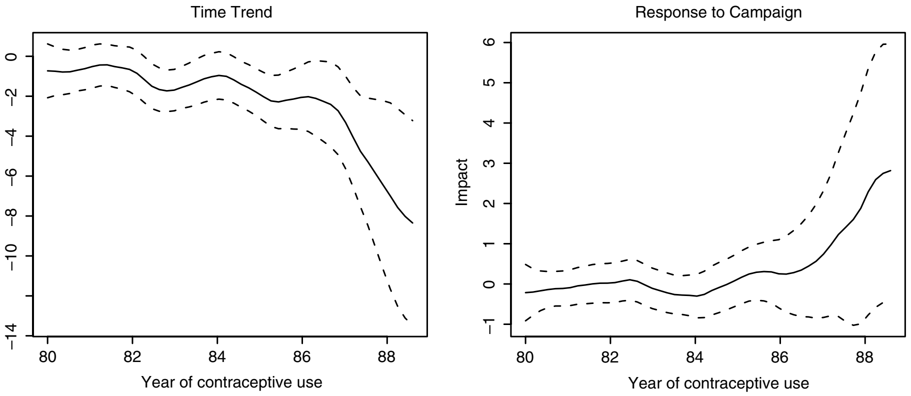

Fig. 5. The solid lines are the estimated impacts of the factors, and dashed lines are the 95% confidence bands.

Whether a woman used contraceptive before does not contribute significantly to the probability of her contraceptive failure either.

Table 6 shows $a_i, i = 2, 3, 4$, is significantly below zero, which indicates the women aged less than 24 are significantly more likely to fail in contraceptive than the women in other age groups. The numerical values of $\hat{a}_2$, $\hat{a}_3$ and $\hat{a}_4$ suggest the women in the age group of 25 to 29 are significantly more likely to fail in contraceptive than the women in the age group of 30 to 34, and the women aged over 35 have smallest failure rate of contraceptive. Table 6 also shows a_{13} is significantly larger than zero, which suggests the women who have record of failure of contraceptive before are significantly more likely to fail in contraceptive than the women who do not.

The proposed estimation method is applied to estimate the functional coefficients $a_1(\cdot)$ and $a_{14}(\cdot)$, and the bootstrap based approach is used to construct their confidence bands as the bootstrap based approach outperforms the asymptotic distribution based approach. The resulting estimates and confidence bands are depicted in Fig. 5. Fig. 5 shows the failure rate of contraceptive in China is decreasing with time in general, and from 1986 to 1988 has seen a very sharp decrease. This is generally attributed to more and more effective contraceptive methods being introduced in China during the years. It is very interesting to see the dynamic pattern of the impact of the motivation. There is little difference on the failure rate of contraceptive between the women self-motivated to use contraceptive and those for response to the campaigns before 1985, which suggests the campaigns to encourage women using contraceptive did have some effects on women's attitude towards contraceptive use before 1985. However, the picture after 1985 is quite different. After 1985, the difference is going up steadily, after 1986 in particular it is going up substantially fast. This suggests the women using contraceptive just for response to the campaigns were more and more likely to fail than those self-motivated after 1985, which indicates the campaigns to encourage women using contraceptive were becoming more and more less effective on women's attitudes towards contraceptive use after 1985.

Acknowledgments

We thank the associate editor and the referees for their very helpful comments. Heng Peng's research is supported by CERG Grants of Hong Kong Research Grant Council (HKBU 201707 and HKBU 201809), FRG Grant from Hong Kong Baptist University FRG/08-09/II-33, and a Grant from National Nature Science Foundation of China (NNSF 10871054).

Appendix. Proof of theorems

Before we present the proofs of the theorems, we first impose some regularity conditions.

(C1) The function $q_2(s, y) < 0$ for $s \in \mathbb{R}$ and y in the range of the response variable.
(C2) The density function of $U, f(u)$ is continuous and positive on the interval $\Omega = [0, 1]$. The elements of the function $\mathbf{a}^{(2)}(\cdot)$ are continuous in $u \in \Omega$.
(C3) The function $f''(u)$, $V(m(u, \mathbf{x}))$, $V'(m(u, \mathbf{x}))$, $V''(m(u, \mathbf{x}))$ and $g'''(m(u, \mathbf{x}))$ are continuous.
(C4) $\Gamma(u) > 0$ for $u \in \Omega$.
(C5) The kernel function $K(\cdot)$ is symmetric density function, and is absolutely continuous on its support set $[-A, A]$.
C5a. $K(A) \neq 0$ or
C5b. $K(A) = 0$, $K(z)$ is absolutely continuous and $K^2(z)$, $(K'(z))^2$ are integrable on the $(-\infty, +\infty)$.
C5c. The function $t^3K(t)$ and $t^3K'(t)$ are bounded and $\int t^4K'(t) < \infty$.
(C6) For an $s > 2$, $\mathrm{E}(|\mathbf{X}|^{2s}|U = u) < \infty$ is continuous and $\mathrm{E}(Y^{2s}|U = u, \mathbf{X} = \mathbf{x}) < \infty$.

Condition (C1) is imposed so that the local likelihood is concave in the parameters, which ensures the uniqueness of the solution. It is satisfied for the canonical exponential family with a canonical link. Conditions (C2) and (C3) imply that $q_1(\cdot, \cdot)$, $q_2(\cdot, \cdot)$, $q_3(\cdot, \cdot)$, $\rho'(\cdot, \cdot)$ and $m'(\cdot)$ are continuous.

To obtain the proof of the theorems, the following lemmas are required.

Lemma A.1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random vectors, where Y_i's are scalar random variables. Assume further that $\mathrm{E}|y|^s < \infty$ and $\sup_x \int |y|^s f(x, y)\mathrm{d}y < \infty$, where f denotes the joint density of (X, Y). Let K be a bounded positive function with a bounded support, satisfying a Lipschitz condition, and D be a compact set, then*

$$\sup_{x\in D} \left| \frac{1}{n} \sum_{i=1}^{n} [K_h(X_i - x)Y_i - \mathrm{E}\{K_h(X_i - x)Y_i\}] \right| = O_p\left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right),$$

provided that $n^{2\varepsilon-1}h \to \infty$ for some $\varepsilon < 1 - s^{-1}$.

This lemma follows immediately from a result in [25].

Next we introduce a useful lemma which will be applied to prove our main result. It is interesting in its own right. Let $(U_1, \xi_1), \ldots, (U_n, xi_n)$ be i.i.d. random samples from (U, ξ). We assume that U and the kernel function $K(\cdot)$ satisfy the above regularity conditions and ξ satisfy

(a) *for an $s > 2$, $\mathrm{E}|\xi|^s < \infty$;*
(b) *the function $r(u)$ is bounded away from zero for $u \in [0, 1]$, and has a bounded first derivative on Ω, where $r(u) = \mathrm{E}(\xi^2|U = u)$;*
(c) *$\sup_x \int |y|^s f(x, y)\mathrm{d}y = c_s < \infty$, where $f(x, y)$ is the joint function of U and ξ.*

Let

$$\mathbf{m}(u) = \frac{1}{\sqrt{nhf(u)r(u)}} \sum_{i=1}^{n} \xi_i K\left(\frac{U_i - u}{h}\right) \quad \text{and} \quad \mathbf{M}(u) = \mathbf{m}(u) - \mathrm{E}\mathbf{m}(u).$$

Lemma A.2. *Under assumptions (a)–(c) and regularity conditions above, if $h = n^{-b}$, for some $0 < b < 1 - 2/s$, we have*

$$P\{(-2\log h)^{1/2}(\nu_0^{-1/2}\|\mathbf{M}\|_\infty - d_n) < x\} \to \exp\{-2\exp(-x)\}$$

where with $\nu_0 = \int K^2(t)\mathrm{d}t$,

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}} \left\{ \log \frac{K^2(A)}{\nu_0 \pi^{1/2}} + \frac{1}{2} \log\log h^{-1} \right\},$$

if assumption (C5a) holds, and

$$d_n = (-2\log h)^{1/2} + \frac{1}{(-2\log h)^{1/2}} \log\left\{ \frac{1}{4\nu_0\pi} \int (K'(t))^2 \mathrm{d}t \right\}$$

if assumption (C5b) is valid.

Lemma A.2 is same as Lemma 1 in [16]. Its proof can be obtained by the technique of the proof for Lemma 3 in [26] and the technique in [27]. Lemma A.1 can be regarded as a corollary of Lemma A.2, except for different technical assumptions.

Let

$$\bar{\eta}(u, U_i, \mathbf{X}_i) = \sum_{j=1}^{p} \left\{ a_j(u) + a_j'(u)(U_i - u) + \cdots + \frac{1}{q!} a_j^{(q)}(U_i - u)^q \right\} X_{ij},$$

$$\hat{\boldsymbol{\beta}}^* = \sqrt{nh} \left\{ (\hat{\mathbf{a}} - \mathbf{a})^T, h(\hat{\mathbf{a}}' - \mathbf{a}')^T, \ldots, h^q(\hat{\mathbf{a}}^{(q)} - \mathbf{a}^{(q)})^T \right\}^T$$

and

$$\Delta(u) = f(u)\Psi_{q+1} \otimes \Gamma(u) \quad \textit{and} \quad \mathbf{Z}_i = (\mathbf{X}_i^T, ((U_i - u)/h)\mathbf{X}_i^T, \ldots, ((U_i - u)/h)^q \mathbf{X}_i^T)^T.$$

Lemma A.3. *Under the regularity conditions given above, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\sup_{u \in [0,1]} \left| \hat{\boldsymbol{\beta}}^* - \Delta^{-1}(u)\mathbf{W}_n \right| = O_p\{h^{q+1} + (nh)^{-1/2}\log^{1/2}(1/h)\}$$

holds, where

$$\mathbf{W}_n = \sqrt{1/nh} \sum_{i=1}^n q_1(\bar{\eta}(u, U_i, \mathbf{X}_i), Y_i)\mathbf{Z}_i K\{(U_i - u)/h\}.$$

Lemma A.3 extend the result of [18] from local linear fitting to q order local polynomial fitting. The proof of this lemma can directly follow steps of the proof for Lemma 2 in Li and Huang (2008) without any change.

If $a_p(u)$ is a constant C_p, let

$$\tilde{\boldsymbol{\beta}}^* = \sqrt{nh}\left\{\tilde{a}_1(u) - a_1(u), \ldots, \tilde{a}_{p-1}(u) - a_{p-1}(u), h(\tilde{a}_1'(u) - a_1'(u)), \ldots, h(\tilde{a}_{p-1}'(u) - a_{p-1}'(u))\right\},$$

$$\mathbf{X}_i^{(1)} = (X_{i1}, \ldots, X_{i(p-1)})^T, \qquad \mathbf{Z}_i^{(1)} = \left(X_{i1}, \ldots, X_{i(p-1)}, \frac{U_i - u}{h}X_{i1}, \ldots, \frac{U_i - u}{h}X_{i(p-1)}\right)^T$$

and

$$\tilde{\eta}(u, U_i, \mathbf{X}_i) = \sum_{j=1}^{p-1}\{a_j(u) + a_j'(u)(U_i - u)\}X_{ij}.$$

We have the following lemma

Lemma A.4. *Under the regularity conditions given above, if $h \to 0$, $nh \to \infty$ as $n \to \infty$ and $\hat{C}_p - C_p = O_p(1/\sqrt{n})$ then*

$$\sup_{u \in [0,1]} \left| \tilde{\boldsymbol{\beta}}^* - \tilde{\Delta}^{-1}(u)\widetilde{\mathbf{W}}_n \right| = o_p(1)$$

where

$$\widetilde{\mathbf{W}}_n = \sqrt{1/nh} \sum_{i=1}^n q_1(\tilde{\eta}(u, U_i, \mathbf{X}_i) + C_p X_{ip}, Y_i)\mathbf{Z}_i^{(1)} K\{(U_i - u)/h\},$$

and

$$\tilde{\Delta}(u) = f(u)\Psi_{q+1} \otimes \tilde{\Gamma}(u), \quad \tilde{\Gamma}(u) = \mathrm{E}\{\rho(U, \mathbf{X})\mathbf{X}^{(1)}\mathbf{X}^{(1)T} | U = u\}.$$

Proof of Lemma A.4. Let $\gamma_n = 1/\sqrt{nh}$, and consider

$$\begin{aligned} l_n^*(\tilde{\boldsymbol{\beta}}^*) &= \sum_{i=1}^n \left[\ell\{g^{-1}(\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}\hat{C}_p + \gamma_n \tilde{\boldsymbol{\beta}}^{*T}\mathbf{Z}_i^{(1)}), Y_i\} - \ell\{g^{-1}(\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}\hat{C}_p), Y_i\}\right] K((U_i - u)/h) \\ &= \mathbf{W}_n^T \tilde{\boldsymbol{\beta}}^* + \frac{1}{2}\tilde{\boldsymbol{\beta}}_n^* \Delta_n \tilde{\boldsymbol{\beta}}^* + R_n^* \end{aligned} \tag{A.1}$$

where

$$\mathbf{W}_n = \gamma_n \sum_{i=1}^n q_1\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}\hat{C}_p, Y_i\}\mathbf{Z}_i^{(1)} K((U_i - u)/h),$$

$$\Delta_n = \gamma_n^2 \sum_{i=1}^n q_2\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}\hat{C}_p, Y_i\}\mathbf{Z}_i^{(1)}\mathbf{Z}_i^{(1)T} K((U_i - u)/h),$$

and

$$R_n^* = \frac{\gamma_n^3}{6} \sum_{i=1}^n q_3(\eta_i + X_{ip}\hat{C}_p, Y_i)(\mathbf{Z}_i^{(1)T}\tilde{\boldsymbol{\beta}}^*)^3 K((U_i - u)/h),$$

η_i is between $\tilde{\eta}(u, U_i, \mathbf{X}_i)$ and $\tilde{\eta}(u, U_i, \mathbf{X}_i) + \gamma_n \tilde{\boldsymbol{\beta}}^{*T}\mathbf{Z}_i^{(1)}$.

Then

$$\mathbf{W}_n - \widetilde{\mathbf{W}}_n = \gamma_n \sum_{i=1}^{n} \left[q_1\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}\hat{C}_p, Y_i\} - q_1\{\tilde{\eta}_i(u, U_i, \mathbf{X}_i) + X_{ip}C_p, Y_i\} \right] \mathbf{Z}_i^{(1)} K((U_i - u)/h),$$

and by Taylor expansion we have that

$$\begin{aligned}\mathbf{W}_n - \widetilde{\mathbf{W}}_n &= \gamma_n \sum_{i=1}^{n} q_2\{\tilde{\eta}_i(u, U_i, \mathbf{X}_i) + X_{ip}C_p, Y_i\}\mathbf{Z}_i^{(1)} X_{ip}(\hat{C}_p - C_p) K((U_i - u_0)/h) \\ &\quad + \gamma_n \sum_{i=1}^{n} q_3\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}C_p^*, Y_i\}\mathbf{Z}_i^{(1)} X_{ip}^2(\hat{C}_p - C_p)^2 K((U_i - u_0)/h).\end{aligned}$$

Since

$$\begin{aligned}&\gamma_n \sum_{i=1}^{n} q_2\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}C_p, Y_i\}\mathbf{Z}_i^{(1)} X_{ip}(\hat{C}_p - C_p) K((U_i - u)/h) \\ &= O_p(nh \cdot n^{-1/2}\gamma_n \mathrm{E}|q_2\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}C_p, Y_i\}\mathbf{Z}_i^{(1)} X_{ip}|) \\ &= O_p(\sqrt{h}) = o_p(1)\end{aligned}$$

and by the similar way, we can also show that

$$\gamma_n \sum_{i=1}^{n} q_3\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}C_p^*, Y_i\}\mathbf{Z}_i^{(1)} X_{ip}^2(\hat{C}_p - C_p)^2 K((U_i - u)/h) = o_p(1),$$

we have

$$\mathbf{W}_n - \widetilde{\mathbf{W}}_n = o_p(1). \tag{A.2}$$

By Taylor expansion and as the same steps shown above for $\mathbf{W}_n$, we can also show that

$$\Delta_n - \tilde{\Delta}_n = o_p(1) \quad \text{and} \quad R_n^* = o_p(1) \tag{A.3}$$

where

$$\tilde{\Delta}_n = \gamma_n^2 \sum_{i=1}^{n} q_2\{\tilde{\eta}(u, U_i, \mathbf{X}_i) + X_{ip}C_p, Y_i\}\mathbf{Z}_i^{(1)}\mathbf{Z}_i^{(1)T} K((U_i - u)/h).$$

Hence by (A.1)–(A.3),

$$\mathbf{W}_n^T \tilde{\boldsymbol{\beta}}^* + \frac{1}{2}\tilde{\boldsymbol{\beta}}_n^{*T} \Delta_n \tilde{\boldsymbol{\beta}}^* - \widetilde{\mathbf{W}}_n^T \tilde{\boldsymbol{\beta}}^* - \frac{1}{2}\tilde{\boldsymbol{\beta}}_n^{*T} \tilde{\Delta}_n \tilde{\boldsymbol{\beta}}^* = o_p(1)$$

and we have

$$l_n^*(\tilde{\boldsymbol{\beta}}^*) = \widetilde{\mathbf{W}}_n^T \tilde{\boldsymbol{\beta}}^* + \frac{1}{2}\tilde{\boldsymbol{\beta}}_n^{*T} \tilde{\Delta}_n \tilde{\boldsymbol{\beta}}^* + o_p(1).$$

Similar as Li and Huang (2008) and [20], by Convexity Lemma (see [28]) and the quadratic approximation lemma (see [19], p. 210), uniformly for $u \in \Omega$, we have

$$\tilde{\boldsymbol{\beta}}^* = \tilde{\Delta}^{-1}(u)\widetilde{\mathbf{W}}_n + o_p(1).$$

So we finished the proof of this lemma. □

Proof of Theorem 1. By Lemma A.3, when $q = 1$ we know that

$$\sup_{u \in \Omega} |\sqrt{nh}(\hat{a}_p(u) - C_p) - e_{p,k}^T \Delta^{-1}(u)\mathbf{W}_n(u)| = O_p\{h^2 + (nh)^{-1/2}\log^{1/2}(1/h)\}$$

where $k = 2p$, and

$$\mathbf{W}_n(u) = \sqrt{1/nh} \sum_{i=1}^{n} q_1(\bar{\eta}(u, U_i, \mathbf{X}_i), Y_i)\mathbf{Z}_i K\{(U_i - u)/h\}.$$

By the equation above, we have

$$\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\hat{a}_p(U_i)-C_p-\frac{e_{p,k}^T}{n\sqrt{nh}}\sum_{i=1}^{n}\Delta^{-1}(U_i)\mathbf{W}_n(U_i)\right| = \sqrt{n}/\sqrt{nh}\cdot O_p(h^2+(nh)^{-1/2}\log^{1/2}(1/h)) = o_p(1).$$

So $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{a}_1(U_i)-C_1=\sqrt{n}(\hat{C}_p-C_p)$ has the same asymptotic distribution with

$$\frac{e_{p,k}^T}{n\sqrt{h}}\sum_{i=1}^{n}\Delta^{-1}(U_i)\mathbf{W}_n(U_i).$$

Next consider the term above, we have

$$\begin{aligned}\frac{\sqrt{n}e_{p,k}^T}{n\sqrt{nh}}\sum_{i=1}^{n}\Delta^{-1}(U_i)\mathbf{W}_n(U_i) &= \frac{\sqrt{n}e_{p,k}^T}{n^2}\sum_{i,j=1}^{n}\Delta^{-1}(U_i)q_1\{\bar{\eta}(U_i,U_j,\mathbf{X}_j),Y_j\}\mathbf{Z}_jK_h\{(U_j-U_i)\}\\ &= \frac{\sqrt{n}e_{p,k}^T}{n^2}\sum_{i\neq j}^{n}\Delta^{-1}(U_i)q_1\{\bar{\eta}(U_i,U_j,\mathbf{X}_j),Y_j\}\mathbf{Z}_jK_h\{(U_j-U_i)\}\\ &\quad+\frac{\sqrt{n}e_{p,k}^T}{n^2}\sum_{i=1}^{n}\Delta^{-1}(U_i)q_1\{\bar{\eta}(U_i,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_jK_h(0)\\ &\hat{=} I_1+I_2.\end{aligned}\tag{A.4}$$

Since $\bar{\eta}(U_i,U_i,\mathbf{X}_i)=\eta(U_i,\mathbf{X}_i)$ and by the property of $q_1(\cdot)$, I_2 can be regarded as the sum of n independent random variables with mean 0. Hence it is not difficult to show that

$$I_2=O_p\left(\frac{1}{nh}\right)=o_p(1).\tag{A.5}$$

Next consider $1/\sqrt{n}I_1$ which can be regarded as a U-statistics with kernel function

$$\psi(i,j)\hat{=}e_{p,k}^T\Delta^{-1}(U_i)q_1\{\bar{\eta}(U_i,U_j,\mathbf{X}_j),Y_j\}\mathbf{Z}_jK_h\{(U_j-U_i)\}+e_{p,k}^T\Delta^{-1}(U_j)q_1\{\bar{\eta}(U_j,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_iK_h\{(U_i-U_j)\}.$$

If $\mathrm{Var}(\mathrm{E}(\psi(i,j)|U_i,X_i,Y_i))<\infty$, then by the center limited theorem of U-statistics (see [29], Theorem 1 in p. 96), I_1 should follow the asymptotical normal distribution with mean $\mathrm{E}\psi(i,j)$ and variance $\mathrm{Var}(\mathrm{E}(\psi(i,j)|U_i,X_i,Y_i))$.

First notice that

$$\begin{aligned}\mathrm{E}(\psi(i,j)|U_i,\mathbf{X}_i,Y_i) &= \mathrm{E}\{e_{p,k}^T\Delta^{-1}(U_i)q_1\{\bar{\eta}(U_i,U_j,\mathbf{X}_j),Y_j\}\mathbf{Z}_jK_h\{(U_j-U_i)\}|U_i,\mathbf{X}_i,Y_i\}\\ &\quad+\mathrm{E}\{e_{p,k}^T\Delta^{-1}(U_j)q_1\{\bar{\eta}(U_j,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_iK_h\{(U_i-U_j)\}|U_i,\mathbf{X}_i,Y_i\}\\ &\hat{=} L_1+L_2.\end{aligned}\tag{A.6}$$

Because $q_k(s,y)$ is linear in y for fixed s, by a Taylor series expansion of $\eta(u,\mathbf{X}_i)$ with respect to u around $|u-u_0|<h$ and (2.2), we have

$$\eta(u,\mathbf{X}_i)=\bar{\eta}(u_0,u,\mathbf{X}_i)+\frac{(u-u_0)^2}{2}\eta_u''(u_0,\mathbf{X}_i)+o(h^2)$$

where $\eta_u''(u,\mathbf{X}_i)=(\partial^2/\partial u^2)\eta(u,\mathbf{X}_i)=\sum_{j=1}^{p}a_j''(u)X_{ij}$, which implies that

$$q_1\{\bar{\eta}(u_0,u,\mathbf{X}_i),m(u,\mathbf{X}_i)\}=\rho(u,\mathbf{X}_i)\frac{(u-u_0)^2}{2}\eta_u''(u_0,\mathbf{X}_i)+o(h^2)\tag{A.7}$$

and

$$q_2(\bar{\eta}(u_0,u,\mathbf{X}_i),m(u,\mathbf{X}_i))=-\rho(u,\mathbf{X}_i)+o(1).\tag{A.8}$$

So by (A.7) we have

$$\begin{aligned}L_1 &= \mathrm{E}\{e_{p,k}^T\Delta^{-1}(U_i)q_1\{\bar{\eta}(U_i,U_j,\mathbf{X}_j),m(U_j,\mathbf{X}_j)\}\mathbf{Z}_jK_h\{(U_j-U_i)\}|U_i,\mathbf{X}_i,Y_i\}\\ &= \mathrm{E}\left\{\left[e_{p,k}^T\Delta^{-1}(U_i)\rho(U_j,\mathbf{X}_j)\frac{(U_j-U_i)^2}{2}\eta_u''(U_i,\mathbf{X}_j)+o(h^2)\right]\mathbf{Z}_jK_h\{(U_j-U_i)\}|U_i,\mathbf{X}_i,Y_i\right\}\\ &= O(h^2).\end{aligned}\tag{A.9}$$

By the similar way, for L_2 we have

$$\begin{aligned} L_2 &= \mathrm{E}\{e_{p,k}^T \Delta^{-1}(U_j) q_1\{\bar{\eta}(U_j, U_i, \mathbf{X}_i), Y_i\} \mathbf{Z}_i K_h\{(U_j - U_i)\} | U_i, \mathbf{X}_i, Y_i\} \\ &= \mathrm{E}\{e_{p,k}^T \Delta^{-1}(U_j) q_1\{\eta(U_i, \mathbf{X}_i), Y_i\} \mathbf{Z}_i K_h\{(U_j - U_i)\} | U_i, \mathbf{X}_i, Y_i\} \\ &\quad + \mathrm{E}\left\{ e_{p,k}^T \Delta^{-1}(U_j) \left\{ \frac{(U_i - U_j)^2}{2} \eta_u''(U_j, \mathbf{X}_i) + o(h^2) \right\} q_2\{\eta_i, Y_i\} \mathbf{Z}_i K_h\{(U_j - U_i)\} | U_i, \mathbf{X}_i, Y_i \right\} \end{aligned}$$

where η_i is between $\bar{\eta}(U_j, U_i, \mathbf{X}_i)$ and $\eta(U_i, \mathbf{X}_i)$. It is easy to know that the second term in the right side of the equation above is $O(h^2)$, so

$$L_2 = \mathrm{E}\{e_{p,k}^T \Delta^{-1}(U_j) q_1\{\eta(U_i, \mathbf{X}_i), Y_i\} \mathbf{Z}_i K_h\{(U_j - U_i)\} | U_i, \mathbf{X}_i, Y_i\} + O(h^2). \tag{A.10}$$

By (A.6), (A.9) and (A.10), we have

$$\begin{aligned} \mathrm{E}(\psi(i,j)|U_i, \mathbf{X}_i, Y_i) &= \mathrm{E}\{e_{p,k}^T \Delta^{-1}(U_j) q_1\{\eta(U_i, \mathbf{X}_i), Y_i\} \mathbf{Z}_i K_h\{(U_j - U_i)\} | U_i, \mathbf{X}_i, Y_i\} + O(h^2) \\ &= e_{p,k}^T \Delta^{-1}(U_i) q_1\{\eta(U_i, \mathbf{X}_i), Y_i\} f(U_i) \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} \otimes \mathbf{X}_i + O(h^2). \end{aligned}$$

Let $\boldsymbol{\mu}_2 = (\mu_0, \mu_1)^T$, it is no difficult to show

$$\begin{aligned} \mathrm{Var}(\mathrm{E}(\psi(i,j)|U_i, X_i, Y_i)) &= \mathrm{var}\left\{ e_{p,k}^T \Delta^{-1}(U_i) q_1\{\eta(U_i, \mathbf{X}_i), Y_i\} f(U_i) \boldsymbol{\mu}_2 \otimes \mathbf{X}_i \right\} + O(h^2) \\ &= e_{p,k}^T \Psi_2^{-1} \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T \Psi_2^{-1} \otimes \mathrm{E}\left\{ \Gamma^{-1}(U_i) q_1^2\{\eta(U_i), Y_i\} \mathbf{X}_i \mathbf{X}_i^T \Gamma^{-1}(U_i) \right\} e_{p,k} + O(h^2), \end{aligned}$$

and hence we have

$$\begin{aligned} \mathrm{Var}(\mathrm{E}(\psi(i,j)|U_i, X_i, Y_i)) &= \mathrm{E}\{e_{p,p}^T \Gamma^{-1}(U) e_{p,p}\} + O(h^2) \\ &= \sigma_1^2 + O(h^2) < \infty. \end{aligned} \tag{A.11}$$

Finally, consider the expectation of $\psi(i,j)$. Similar as L_1 by (A.7) it is easy to show that

$$\begin{aligned} \mathrm{E}\psi(i,j) &= 2\mathrm{E}\left\{ \mathrm{E}\{e_{p,k}^T \Delta^{-1}(U_i) q_1\{\bar{\eta}(U_i, U_j, \mathbf{X}_j), m(U_j, \mathbf{X}_j)\} \mathbf{Z}_j K_h\{(U_j - U_i)\} | U_i, \mathbf{X}_i, Y_i\} \right\} \\ &= \mathrm{E}\left\{ e_{p,k}^T \Delta^{-1}(U_i) \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \otimes \Gamma(U_i) f(U_i) \mathbf{a}^{(2)}(U_i) \right\} + O(h^3) \\ &= O(h^3). \end{aligned} \tag{A.12}$$

According to (A.11), we know that $\mathrm{Var}(\mathrm{E}(\psi(i,j)|U_i, X_i, Y_i)) < \infty$, so by the central limited theorem of U-statistics (see [29]), $\frac{e_{1,k}^T}{n\sqrt{h}} \sum_{i=1}^n \Delta^{-1}(u_i) \mathbf{W}_n(u_i)$ should be following asymptotical normal distribution with mean $O(h^3)$ and variance $\sigma_1^2 + O(h^2)$. On the other hand because

$$O(\sqrt{n}h^3) = o(1) \quad \text{and} \quad \frac{\sigma_1^2 + O(h^2)}{\sigma_1^2} = 1 + o(1),$$

it is no difficult to show that $\frac{e_{1,k}^T}{n\sqrt{h}} \sum_{i=1}^n \Delta^{-1}(u_i) W_n(u_i)$ is also following the asymptotical normal with asymptotical mean 0 and variance σ_1^2, so does $\sqrt{n}(\hat{C}_1 - C_1)$. We finished the proof of Theorem 1. $\square$

Proof of Theorem 2. By Lemma A.3 and the definition of $\hat{\boldsymbol{\beta}}^*$, it is obviously that

$$\begin{aligned} \sup_{u \in [0,1]} \sqrt{nh} |\hat{a}_p(u) - a_p(u) - \mathrm{bias}(\hat{a}_p(u)|\mathcal{D})| &= \sup_{u \in [0,1]} \left| e_{p,k}^T \left(\hat{\boldsymbol{\beta}}^* - \mathrm{E}\{\hat{\boldsymbol{\beta}}^* | \mathcal{D}\} \right) \right| \\ &= \sup_{u \in [0,1]} \left| e_{p,k}^T \left(\Delta^{-1}(u) \mathbf{W}_n - \Delta^{-1}(u) \mathrm{E}\{\mathbf{W}_n | \mathcal{D}\} \right) \right| + O_p\{h^2 + (nh)^{-1/2} \log^{1/2}(1/h)\}. \end{aligned} \tag{A.13}$$

On the other hand, note that $q_1(s, y)$ is linear in y for fixed s,

$$\mathrm{E}\{\mathbf{W}_n | \mathcal{D}\} = \frac{1}{\sqrt{nh}} \sum_{i=1}^n q_1\{\bar{\eta}(u, U_i, \mathbf{X}_i), m(U_i, \mathbf{X}_i)\} \mathbf{Z}_i K\{(U_i - u)/h\}, \tag{A.14}$$

$$\Delta(u) = f(u) \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma(u)$$

and $\mu_1 = 0$ for symmetric density function $K(\cdot)$, we can define

$$\begin{aligned} I &= \sqrt{f(u)} e_{p,k}^T \left(\Delta^{-1}(u)\mathbf{W}_n - \Delta^{-1}(u)\mathrm{E}\{\mathbf{W}_n|\mathcal{D}\} \right) \\ &= \frac{1}{\sqrt{nhf(u)}} \sum_{i=1}^{n} \xi_i K_1\{(U_i - u)/h\} \end{aligned} \tag{A.15}$$

where

$$\xi_i = e_{p,p}^T \Gamma^{-1}(u)(X_{i1}, \ldots, X_{ip})^T (q_1\{\bar{\eta}(u, U_i, \mathbf{X}_i), Y_i\} - q_1\{\bar{\eta}(u, U_i, \mathbf{X}_i), m(U_i, \mathbf{X}_i)\}).$$

According to the results of Cai et al. [20]

$$\begin{aligned} r(u) &= \mathrm{E}(\xi_i^2 | U = u) \\ &= e_{p,p}^T \Gamma^{-1}(u) \mathrm{E}\{q_1^2\{\bar{\eta}(u, U, \mathbf{X}), Y\}\mathbf{X}\mathbf{X}^T | U = u\} \Gamma^{-1}(u) e_{p,p}. \end{aligned} \tag{A.16}$$

By the definition of $q_1(\cdot)$, it is not difficult to show

$$\mathrm{E}\{q_1^2\{\bar{\eta}(u, U, \mathbf{X}), Y\}\mathbf{X}\mathbf{X}^T | U = u\} = \Gamma(u)$$

and so

$$r(u) = e_{p,p}^T \Gamma(u) e_{p,p} = r_p(u). \tag{A.17}$$

Apply Lemma A.2 to I and by (A.13)–(A.17) we have

$$\begin{aligned} &P\left\{ (-2\log h)^{1/2} \left(\nu_{1,0}^{-1/2} \sup_{u\in[0,1]} \left| (nhr_p^{-1}(u)f(u))^{1/2} \left(\hat{a}_p(u) - a_p(u) - \mathrm{bias}(\hat{a}_p(u)|\mathcal{D}) \right) \right| - d_{v,n} \right) < x \right\} \\ &\quad \to \exp\{-2\exp(-x)\}. \end{aligned} \tag{A.18}$$

On the other hand, by a Taylor series expansion of $\eta(u, x)$ with respect to u around $|u - u_0| < h$ and (2.2), we have

$$\eta(u, \mathbf{x}) = \bar{\eta}(u_0, u, \mathbf{x}) + \frac{(u - u_0)^2}{2} \eta_u''(u_0, \mathbf{x}) + O(h^3)$$

where $\eta_u''(u, \mathbf{x}) = (\partial^2/\partial u)\eta(u, \mathbf{x}) = \sum_{j=1}^{p} a_j''(u)x_j$, which implies that

$$q_1\{\bar{\eta}(u_0, u, \mathbf{x}), m(u, \mathbf{x})\} = \rho(u, \mathbf{x}) \frac{(u - u_0)^2}{2} \eta_u''(u_0, \mathbf{x}) + O(h^3).$$

By Lemma A.1, condition (C5) and (A.13), we have

$$\begin{aligned} \mathrm{E}\{\mathbf{W}_n|\mathcal{D}\} &= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} q_1\{\bar{\eta}(u, U_i, \mathbf{X}_i), m(U_i, \mathbf{X}_i)\} \mathbf{Z}_i K\{(U_i - u)/h\} \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \left\{ \rho(U_i, \mathbf{X}_i) \frac{(U_i - u)^2}{2} \eta_u''(u, \mathbf{X}_i) + O(h^3) \right\} \mathbf{Z}_i K\{(U_i - u)/h\} \\ &= \frac{\sqrt{nh} h^2 f(u)}{2} \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \otimes \Gamma(u) \mathbf{a}''(u)(1 + O(h)) \end{aligned}$$

uniformly for u. By (A.14) we have

$$\begin{aligned} \sqrt{nh} \cdot \mathrm{bias}(\hat{a}_p(u)|\mathcal{D}) &= e_{p,k}^T \Delta^{-1}(u) \mathrm{E}\{\mathbf{W}_n|\mathcal{D}\} \\ &= \frac{\sqrt{nh} h^2}{2} e_{p,k}^T \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}^{-1} \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \otimes \mathbf{I}_k \mathbf{a}''(u)(1 + O(h)) \\ &= \frac{\sqrt{nh^5} \mu_2}{2} a_p''(u)(1 + O(h)). \end{aligned}$$

Hence, uniformly for u, we have

$$(-2\log h)^{1/2} \nu_{1,0}^{-1/2} \sup_{u\in[0,1]} \left| (nhr_p^{-1}(u)f(u))^{1/2} \cdot \left(\mathrm{bias}(\hat{a}_p(u)|\mathcal{D}) - \frac{h^2 \mu_2}{2} a_p''(u) \right) \right| = o_p(1). \tag{A.19}$$

By (A.18) and (A.19), we can replace $\mathrm{bias}(\hat{a}_p(u))$ by $\frac{h^2\mu_2}{2}a''_p(u)$, and (3.1)

$$P\left\{(-2\log h)^{1/2}(\nu_{1,0}^{-1/2}\sup_{u\in[0,1]}\left|(nhr_p^{-1}(u)f(u))^{1/2}\left(\hat{a}_p(u)-a_p(u)-\frac{h^2\mu_2}{2}a''_p(u)\right)\right|-d_{v,n})<x\right\}\to\exp\{-2\exp(-x)\}$$

has been proved.

Next consider the difference between $\widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D})$ and $2^{-1}h^2\mu_2a''_p(u)$. By Lemma A.3, we have

$$\begin{aligned}(-2\log h)^{1/2}(nh)^{1/2}\left\{\widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D})-2^{-1}h^2\mu_2a''_p(u)\right\} &= (-2\log h)^{1/2}(nh)^{1/2}h^2\left\{h_*^2+(nh_*^5)^{-1/2}\log^{1/2}(1/h_*)\right\}\\ &= o(1)\end{aligned}\tag{A.20}$$

uniformly for u, where $h_*=O(n^{-1/9})$. So by (A.20), we have

$$\begin{aligned}&P\left\{(-2\log h)^{1/2}\left(\nu_{1,0}^{-1/2}\sup_{u\in[0,1]}\left|(nhr_p^{-1}(u)f(u))^{1/2}(\hat{a}_p(u)-a_p(u)-\widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D}))\right|-d_{v,n}\right)<x\right\}\\ &\quad\to\exp\{-2\exp(-x)\}.\end{aligned}\tag{A.21}$$

According to (A.21), to finish the proof we only need prove that

$$\sup_{u\in[0,1]}\left|nh\,\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})-\nu_{1,0}r_p(u)f^{-1}(u)\right|=o_p(1).\tag{A.22}$$

Let $\hat{\eta}(u,U_i,\mathbf{X}_i)=\sum_{j=1}^p\{\hat{a}_j+\hat{b}_j(U_i-u)/h\}X_{ij}$, and notice that

$$\mathbf{Z}_i\mathbf{Z}_i^T=\left(1,\frac{U_i-u}{h}\right)^T\left(1,\frac{U_i-u}{h}\right)\otimes(\mathbf{X}_i\mathbf{X}_i^T),$$

we have

$$nh\cdot\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})=e_{p,k}\Gamma^{-1}(\hat{\mathbf{a}},\hat{\mathbf{b}})\Lambda(\hat{\mathbf{a}},\hat{\mathbf{b}})\Gamma^{-1}(\hat{\mathbf{a}},\hat{\mathbf{b}})e_{p,k}^T\tag{A.23}$$

where

$$\Gamma(\hat{\mathbf{a}},\hat{\mathbf{b}})=-\frac{1}{n}\sum_{i=1}^n q_2\{\hat{\eta}(u,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_i\mathbf{Z}_i^TK_h(U_i-u),\tag{A.24}$$

and

$$\Lambda(\hat{\mathbf{a}},\hat{\mathbf{b}})=\frac{h}{n}\sum_{j=1}^n q_1^2\{\hat{\eta}(u,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_i\mathbf{Z}_i^TK_h^2(U_i-u).\tag{A.25}$$

Define

$$\hat{\boldsymbol{\beta}}=\gamma_n(\hat{a}_1-a_1(u),\ldots,\hat{a}_p-a_p(u),\hat{b}_1-ha'_1(u_0),\ldots,\hat{b}_p-ha'_p(u))^T$$

where $\gamma_n=(nh)^{-1/2}$. It can be easily seen that

$$\sum_{j=1}^p\{\hat{a}_j+\hat{b}_j(U_i-u)/h\}X_{ij}=\bar{\eta}(u,U_i,\mathbf{X}_i)+\gamma_n\hat{\boldsymbol{\beta}}^T\mathbf{Z}_i.$$

Hence by Taylor expansion, for (A.24) we have

$$\Gamma(\hat{\mathbf{a}},\hat{\mathbf{b}})=-\frac{1}{n}\sum_{i=1}^n q_2\{\bar{\eta}(u,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_i\mathbf{Z}_i^TK_h(U_i-u)-\frac{\gamma_n}{n}\sum_{i=1}^n\hat{\boldsymbol{\beta}}^T\mathbf{Z}_iq_3\{\eta_i,Y_i\}\mathbf{Z}_i\mathbf{Z}_i^TK_h(U_i-u)\tag{A.26}$$

where η_i is between $\bar{\eta}(u,U_i,\mathbf{X}_i)$ and $\bar{\eta}(u,U_i,\mathbf{X}_i)+\gamma_n\hat{\boldsymbol{\beta}}^T\mathbf{Z}_i$.

By Lemmas A.1 and A.3 and Condition (C6), the second term of the (A.26) is bounded

$$\frac{\gamma_n}{n}\sum_{i=1}^n\hat{\boldsymbol{\beta}}^T\mathbf{Z}_iq_3\{\eta_i,Y_i\}\mathbf{Z}_i\mathbf{Z}_i^TK_h(U_i-u)=O_p(\gamma_n\mathrm{E}|q_3(\eta_1,Y_1)\mathbf{X}_1^3K_h(U_1-u)|)=O_p(\gamma_n)\tag{A.27}$$

uniformly for u in [0, 1]. Similar as [20], and by Lemma A.1 it is easy to show

$$-\frac{1}{n}\sum_{i=1}^n q_2\{\bar{\eta}(u,U_i,\mathbf{X}_i),Y_i\}\mathbf{Z}_i\mathbf{Z}_i^TK_h(U_i-u)=-\Delta(u)+O_p\left(h^2+\left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right)\tag{A.28}$$

uniformly for u in $[0, 1]$, where

$$\Delta(u) = f(u)\begin{pmatrix}\mu_0 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}\otimes\Gamma(u).$$

So by (A.26)–(A.28), uniformly for u in $[0, 1]$, we have

$$\Gamma(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = -\Delta(u) + o_p(1). \tag{A.29}$$

With the similar way, for (A.25) we can show that

$$\Lambda(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \frac{h}{n}\sum_{j=1}^{n} q_1^2\{\bar{\eta}_i(u, U_i, \mathbf{X}_i), Y_i\}\mathbf{Z}_i\mathbf{Z}_i^T K_h^2(U_i - u_0) + \frac{h\gamma_n}{n}\sum_{j=1}^{n} 2\hat{\boldsymbol{\beta}}^T\mathbf{Z}_i q_1\{\eta_i^*, Y_i\}q_2\{\eta_i^*, Y_i\}\mathbf{Z}_i\mathbf{Z}_i^T K_h^2(U_i - u) \tag{A.30}$$

where η_i^* is between $\bar{\eta}(u, U_i, \mathbf{X}_i)$ and $\bar{\eta}(u, U_i, \mathbf{X}_i) + \gamma_n\hat{\boldsymbol{\beta}}^T\mathbf{Z}_i$.

For the second term of (A.30), we have

$$\frac{h\gamma_n}{n}\sum_{j=1}^{n} 2\hat{\boldsymbol{\beta}}^T\mathbf{Z}_i q_1\{\eta_i^*, Y_i\}q_2\{\eta_i^*, Y_i\}\mathbf{Z}_i\mathbf{Z}_i^T K_h^2(U_i - u) = O_p(\gamma_n). \tag{A.31}$$

For the first term of (A.30), we have

$$\begin{aligned}\frac{h}{n}\sum_{j=1}^{n} q_1^2\{\bar{\eta}_i(u, U_i, \mathbf{X}_i), Y_i\}\mathbf{Z}_i\mathbf{Z}_i^T K_h^2(U_i - u) &= \Lambda(u) + O_p\left(h^2 + \left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right)\\ &= \Lambda(u) + o_p(1)\end{aligned} \tag{A.32}$$

where

$$\Lambda(u) = f(u)\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}\otimes\Gamma(u).$$

By (A.29)–(A.32) and simple computation, we have

$$e_{p,k}\Gamma^{-1}(\hat{\mathbf{a}}, \hat{\mathbf{b}})\Lambda(\hat{\mathbf{a}}, \hat{\mathbf{b}})\Gamma^{-1}(\hat{\mathbf{a}}, \hat{\mathbf{b}})e_{1,k}^T = e_{p,k}\Delta^{-1}(u)\Lambda(u)\Delta^{-1}(u)e_{p,k}^T + o_p(1) = \frac{r_p(u)\nu_{1,0}}{f(u)} + o_p(1)$$

uniformly for u in $[0, 1]$, so we proved (A.22). With (A.20) and (A.22), it is not difficult to see that (3.2) has been proved, and hence we finished the proof of Theorem 2. □

Proof of Theorem 3. Since $a_p(u) = C_p$ is a constant, by Theorem 2 we have

$$\begin{aligned}&P\left\{(-2\log h)^{1/2}\left(\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}(\hat{a}_p(u) - C_p - \widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D}))\right| - d_{v,n}\right) < x\right\}\\ &\quad\to \exp\{-2\exp(-x)\}\end{aligned} \tag{A.33}$$

and

$$\sup_{u\in[0,1]}|\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}| = O_p\left(\frac{\log^{1/2}(1/h)}{\sqrt{nh}}\right). \tag{A.34}$$

Next by Theorem 1, we know that

$$\hat{C}_p - C_p = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.35}$$

By (A.34) and (A.35), it is not difficult to show that

$$\begin{aligned}&(-2\log h)^{1/2}\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}\left(\hat{a}_p(u) - \hat{C}_1 - \widehat{\mathrm{bias}}(a_p(u)|\mathcal{D})\right)\right|\\ &\quad= (-2\log h)^{1/2}\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}\left(\hat{a}_p(u) - C_p - \widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D}) + C_p - \hat{C}_p\right)\right|\\ &\quad= (-2\log h)^{1/2}\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}(\hat{a}_p(u) - C_p - \widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D}))\right| + o_p(1).\end{aligned}$$

So

$$(-2\log h)^{1/2}\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}\left(\hat{a}_p(u)-\hat{C}_p-\widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D})\right)\right|$$

has the same asymptotic distribution with

$$(-2\log h)^{1/2}\sup_{u\in[0,1]}\left|\frac{1}{\{\widehat{\mathrm{var}}(\hat{a}_p(u)|\mathcal{D})\}^{\frac{1}{2}}}\left(\hat{a}_p(u)-C_p-\widehat{\mathrm{bias}}(\hat{a}_p(u)|\mathcal{D})\right)\right|,$$

and hence by (A.33) Theorem 3 has been proved. □

Proof of Theorem 4. According to the definition of $\mathcal{R}$, we have

$$\mathcal{R}=I_1+I_2+I_3 \tag{A.36}$$

where

$$I_1=\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\mathbf{X}_i^T\hat{\mathbf{a}}(U_i)\right\},Y_i\right]-\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\sum_{j=1}^{p-1}X_{ij}a_j(U_i)+X_{ip}C_p\right\},Y_i\right],$$

$$I_2=-\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\sum_{j=1}^{p-1}X_{ij}\tilde{a}_j(U_i)+X_{ip}\hat{C}_p\right\},Y_i\right]+\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\sum_{j=1}^{p-1}X_{ij}\tilde{a}_j(U_i)+X_{ip}C_p\right\},Y_i\right]$$

and

$$I_3=-\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\sum_{j=1}^{p-1}X_{ij}\tilde{a}_j(U_i)+X_{ip}C_p\right\},Y_i\right]+\sum_{i=1}^{n}\ell\left[g^{-1}\left\{\sum_{j=1}^{p-1}X_{ij}a_j(U_i)+X_{ip}C_p\right\},Y_i\right].$$

For I_2, notice that $\hat{C}_p-C_p=O_p(1/\sqrt{n})$, by tedious computation, it is not difficult to show that $I_2=o(1/\sqrt{h})$.

Next we consider I_1. Let $\Gamma(u)=f(u)\mathrm{E}[\rho(U,\mathbf{X})\mathbf{X}\mathbf{X}^T|U=u]$, $\varepsilon_i=q_1(g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i)$ and $r_n=1/\sqrt{nh}$. Using Taylor expansion we have

$$\begin{aligned}I_1&=\sum_{i=1}^{n}q_1\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i\right\}\mathbf{X}_i^T(\hat{\mathbf{a}}(U_i)-\mathbf{a}(U_i))+\frac{1}{2}\sum_{i=1}^{n}q_2\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i\right\}\\&\quad\times(\hat{\mathbf{a}}(U_i)-\mathbf{a}(U_i))^T\mathbf{X}_i\mathbf{X}_i^T(\hat{\mathbf{a}}(U_i)-\mathbf{a}(U_i))+nO_p\left(\left(\frac{-\log h}{\sqrt{nh}}\right)^3+h^6\right)\\&=T_1+T_2+o_p(1/\sqrt{h})\end{aligned}$$

where

$$\begin{aligned}T_1&=r_n^2\sum_{i=1}^{n}q_1\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i\right\}\mathbf{X}_i^T\Gamma^{-1}(U_i)\sum_{j=1}^{n}q_1g^{-1}(\bar{\eta}(U_i,U_j,\mathbf{X}_j),Y_j)\mathbf{X}_jK((U_j-U_i)/h)\\&=r_n^2\sum_{i=1}^{n}q_1\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(u_i)),Y_i\right\}\mathbf{X}_i^T\Gamma^{-1}(U_i)\sum_{j=1}^{n}q_1g^{-1}(\mathbf{X}_j^T\mathbf{a}(U_j),Y_j)\mathbf{X}_jK((U_j-U_i)/h)+R_{n1g}\end{aligned} \tag{A.37}$$

and

$$\begin{aligned}T_2&=\frac{1}{2}\sum_{i=1}^{n}q_2\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i\right\}(\hat{\mathbf{a}}(U_i)-\mathbf{a}(U_i))^T\mathbf{X}_i\mathbf{X}_i^T(\hat{\mathbf{a}}(U_i)-\mathbf{a}(U_i))\\&=\frac{r_n^4}{2}\sum_{i=1}^{n}\sum_{j,k=1}^{n}q_2\left(\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i\right\}\right)\mathbf{X}_j^T\Gamma^{-1}(U_i)q_1g^{-1}(\bar{\eta}(U_i,U_j,\mathbf{X}_j),Y_j)K((U_j-U_i)/h)\\&\quad\times\mathbf{X}_i\mathbf{X}_i^T\Gamma^{-1}(U_i)q_1(g^{-1}(\bar{\eta}(U_i,U_k,\mathbf{X}_k)),Y_k)\mathbf{X}_kK((U_k-U_i)/h)\\&=\frac{r_n^4}{2}\sum_{i=1}^{n}\sum_{j,k=1}^{n}q_2\left(\left\{g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)),Y_i\right\}\right)\mathbf{X}_j^T\Gamma^{-1}(U_i)q_1(g^{-1}(\mathbf{X}_j^T\mathbf{a}(U_j)),Y_j)K((U_j-U_i)/h)\\&\quad\times\mathbf{X}_i\mathbf{X}_i^T\Gamma^{-1}(U_i)q_1(g^{-1}(\mathbf{X}_k^T\mathbf{a}(u_k)),Y_k)\mathbf{X}_kK((U_k-U_i)/h)+R_{n2g}+R_{n3g}.\end{aligned} \tag{A.38}$$

By (A.37) and (A.38), I_1 can be written as

$$I_1 = r_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j \mathbf{X}_j^T \Gamma^{-1}(U_i) \mathbf{X}_i K((U_j - U_i)/h) + R_{n1g} + \frac{r_n^4}{2} \sum_{i=1}^{n} \sum_{j,k=1}^{n} q_2(g^{-1}(\mathbf{X}_i^T \mathbf{a}(u_i)), Y_i) \varepsilon_j \varepsilon_k$$
$$\times \Gamma^{-1}(U_i) \mathbf{X}_j^T \mathbf{X}_i \mathbf{X}_i^T \Gamma^{-1}(U_i) \mathbf{X}_k K((U_j - U_i)/h) K((U_k - U_i)/h) + R_{n2g} + R_{n3g} \quad \text{(A.39)}$$

where

$$R_{n1g} = r_n^2 \sum_{k=1}^{n} \varepsilon_k H_n(U_k) \mathbf{X}_k$$
$$R_{n2g} = r_n^2 \sum_{k=1}^{n} \sum_{i=1}^{n} \varepsilon_i \mathbf{X}_i \Gamma^{-1}(U_k) \mathbf{X}_k \mathbf{X}_k^T H_n(U_k)$$
$$R_{n3g} = -\frac{r_n^4}{2} \sum_{i=1}^{n} q_2(g^{-1}(\mathbf{X}_i^T \mathbf{a}(U_i)), Y_i) H_n(U_k)^T \mathbf{X}_k \mathbf{X}_k H_n(U_k)$$

and

$$H_n(u) = r_n^2 \Gamma^{-1}(u) \sum_{i=1}^{n} [q_1(\bar{\eta}(u, U_i, \mathbf{X}_i), Y_i) - q_1(\mathbf{X}_i \mathbf{a}(U_i), Y_i)] \mathbf{X}_i K((U_i - u)/h)(1 + o_p(1)),$$

and $o_p(1)$ is uniform with respect to u.

Recall that $\Gamma(u) = f(u) \mathrm{E}\{\rho(U, \mathbf{X}) \mathbf{X}\mathbf{X}^T | U = u\}$. Let $\mathbf{X}_i^{(1)} = (X_{i1}, X_{i2}, \ldots, X_{i(p-1)})^T$ and $\Gamma_{11}(u) = f(u) \mathrm{E}\{\rho(U, \mathbf{X}) \mathbf{X}^{(1)} \mathbf{X}^{(1)T} | U = u\}$. Write

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \quad \text{and} \quad \Gamma_{22,1} = \Gamma_{22} - \Gamma_{21} \Gamma_{11}^{-1} \Gamma_{12}.$$

By the similar way, we can show that

$$I_3 = -r_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j \mathbf{X}_j^{(1)T} \Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)} K((U_j - U_i)/h) + R_{n1g}^* - \frac{r_n^4}{2} \sum_{i=1}^{n} \sum_{j,k=1}^{n} q_2(g^{-1}(\mathbf{X}_i^T \mathbf{a}(U_i)), Y_i) \varepsilon_j \varepsilon_k \mathbf{X}_j^{(1)T}$$
$$\Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)} \mathbf{X}_i^{(1)T} \Gamma_{11}^{-1}(U_i) \mathbf{X}_k^{(1)} K((U_j - U_i)/h) K((U_k - U_i)/h) + R_{n2g}^* + R_{n3g}^* \quad \text{(A.40)}$$

where R_{n1g}^*, R_{n2g}^* and R_{n3g}^* is defined by replacing $\mathbf{X}$ and Γ by $\mathbf{X}^{(1)}$ and Γ_{11} in R_{n1g}, R_{n2g} and R_{n3g}.

So by (A.39) and (A.40), (A.36) can be rewritten as

$$\begin{aligned}
\mathcal{R} &= r_n^2 \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j (X_{ip} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)})^T \Gamma_{22,1}^{-1}(U_i)(X_{ip} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)}) K((U_j - U_i)/h) \\
&+ \frac{r_n^4}{2} \sum_{j,k=1}^{n} \varepsilon_j \varepsilon_k \sum_{i=1}^{n} q_2(g^{-1}(\mathbf{X}_i^T \mathbf{a}(u_i)), Y_i)(X_{jp} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_j^{(1)})^T \Gamma_{22,1}^{-1}(U_i) \\
&\times (X_{ip} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)})(X_{ip} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_j^{(1)})^T \Gamma_{22,1}^{-1}(U_i) \\
&\times (X_{kp} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_k^{(1)}) K((U_j - U_i)/h) K((U_k - U_i)/h) \\
&+ R_{n4} + R_{n5} + o_p(h^{-1/2}) + R_{n1g} + R_{n2g} + R_{n3g} - R_{n1g}^* - R_{n2g}^* - R_{n3g}^*
\end{aligned}$$

where

$$\begin{aligned}
R_{n4} &= \frac{r_n^4}{2} \sum_{j,k=1}^{n} \varepsilon_j \varepsilon_k \sum_{i=1}^{n} q_2(g^{-1}(\mathbf{X}_i^T \mathbf{a}(U_i)), Y_i)(X_{jp} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_j^{(1)})^T \Gamma_{22,1}^{-1}(U_i) \\
&\times (X_{ip} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)T} \Gamma_{22,1}^{-1}(U_i) \mathbf{X}_k^{(1)} K((U_j - U_i)/h) K((U_k - U_i)/h)
\end{aligned}$$

and

$$\begin{aligned}
R_{n5} &= \frac{r_n^4}{2} \sum_{j,k=1}^{n} \varepsilon_j \varepsilon_k \sum_{i=1}^{n} q_2(g^{-1}(\mathbf{X}_i^T \mathbf{a}(U_i)), Y_i)(X_{kp} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_k^{(1)})^T \Gamma_{22,1}^{-1}(U_i) \\
&\times (X_{ip} - \Gamma_{21}(U_i) \Gamma_{11}^{-1}(U_i) \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)T} \Gamma_{22,1}^{-1}(U_i) \mathbf{X}_j^{(1)} K((U_j - U_i)/h) K((U_k - U_i)/h).
\end{aligned}$$

By simple calculation, it can easily to show that as $nh^2 \to \infty$

$$\mathrm{E}R_{n4}^2 = O\left(\frac{1}{n^2h^4}\right) = o\left(\frac{1}{h}\right) \quad \text{and} \quad \mathrm{E}R_{n4}^2 = O\left(\frac{1}{n^2h^4}\right) = o\left(\frac{1}{h}\right)$$

which yields $R_{n4} = o_p(h^{-1/2})$ and $R_{n5} = o_p(h^{-1/2})$.

On the other hand, similar as the proof of Fan et al. [23], when $nh^6 \to 0$ and $nh^2/\log h \to \infty$, we have

$$R_{n1g} + R_{n2g} + R_{n3g} - R_{n1g}^* - R_{n2g}^* - R_{n3g}^* = o_p(h^{-1/2})$$

and hence

$$\begin{aligned}
\mathcal{R} &= r_n^2 \sum_{i,j=1}^n \varepsilon_i\varepsilon_j (X_{ip} - \Gamma_{21}(U_i)\Gamma_{11}^{-1}(U_i)\mathbf{X}_i^{(1)})^T \Gamma_{22,1}^{-1}(U_i)(X_{ip} - \Gamma_{21}(U_i)\Gamma_{11}^{-1}(U_i)\mathbf{X}_i^{(1)})K((U_j - U_i)/h) \\
&\quad + \frac{r_n^4}{2} \sum_{j,k=1}^n \varepsilon_j\varepsilon_k \sum_{i=1}^n q_2(g^{-1}(\mathbf{X}_i^T\mathbf{a}(U_i)), Y_i)(X_{jp} - \Gamma_{21}(U_i)\Gamma_{11}^{-1}(U_i)\mathbf{X}_j^{(1)})^T \Gamma_{22,1}^{-1}(U_i) \\
&\quad \times (X_{ip} - \Gamma_{21}(U_i)\Gamma_{11}^{-1}(U_i)\mathbf{X}_i^{(1)})(X_{ip} - \Gamma_{21}(U_i)\Gamma_{11}^{-1}(U_i)\mathbf{X}_j^{(1)})^T \Gamma_{22,1}^{-1}(U_i) \\
&\quad \times (X_{kp} - \Gamma_{21}(U_i)\Gamma_{11}^{-1}(U_i)\mathbf{X}_k^{(1)})K((U_j - U_i)/h)K((U_k - U_i)/h) + o_p(h^{-1/2}).
\end{aligned}$$

Notice that

$$\mathrm{E}[\varepsilon_i|\mathbf{X}_i, U_i] = 0, \qquad \mathrm{E}[\varepsilon_i^2|\mathbf{X}_i, U_i] = -\mathrm{E}[q_2(g^{-1}(\mathbf{X}_i^T\mathbf{a}^T(U_i), Y_i))|\mathbf{X}_i, U_i],$$

then the remain proof can follow the proof of Theorems 5 and 6 in [23]. □

References

[1] L. Breiman, J.H. Friedman, Estimating optimal transformations for multiple regression and correlation (with discussion), J. Amer. Statist. Assoc. 80 (1985) 580–619.
[2] T.J. Hastie, R. Tibshirani, Generalized Additive Models, Chapman and Hall, London, 1990.
[3] J. Fan, W. Härdle, E. Mammen, Direct estimation of additive and linear components for high dimensional data, Ann. Statist. 26 (1998) 943–971.
[4] J. Jiang, H. Zhou, Additive hazards regression with auxiliary covariates, Biometrika 94 (2007) 359–369.
[5] J. Jiang, H. Zhou, X. Jiang, J. Peng, Generalized likelihood ratio tests for the structures of semiparametric additive models, Canad. J. Statist. 35 (2007) 381–398.
[6] J.H. Friedman, Multivariate adaptive regression splines (with discussion), Ann. Statist. 19 (1991) 1–141.
[7] C.J. Stone, M. Hansen, C. Kooperberg, Y.K. Truong, Polynomial splines and their tensor products in extended linear modeling, Ann. Statist. 25 (1997) 1371–1470.
[8] W. Härdle, T.M. Stoker, Investigating smooth multiple regression by the method of average derivatives, J. Amer. Statist. Assoc. 84 (1989) 986–995.
[9] K.-C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 316–342.
[10] G. Wahba, Partial spline models for semiparametric estimation of functions of several variables, in: Statistical Analysis of Time Series, in: Proceedings of the Japan US Joint Seminar, Tokyo, Institute of Statistical Mathematics, Tokyo, 1984, pp. 319–329.
[11] P.J. Green, B.W. Silverman, Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman and Hall, London, 1994.
[12] R.J. Carroll, J. Fan, I. Gijbels, M.P. Wand, Generalized partially linear single-index models, J. Amer. Statist. Assoc. 92 (1997) 477–489.
[13] W.S. Cleveland, E. Grosse, W.M. Shyu, Local regression models, in: J.M. Chambers, T.J. Hastie (Eds.), Statistical Models in S, Wadsworth & Brooks, Pacific Grove, 1991, pp. 309–376.
[14] T.J. Hastie, R.J. Tibshirani, Varying-coefficient models, J. R. Statist. Soc. Ser. B 55 (1993) 757–796.
[15] Y. Xia, W.K. Li, On the estimation and testing of functional-coefficient linear models, Statist. Sinica 9 (1999) 735–758.
[16] J. Fan, W. Zhang, Statistical estimation in varying coefficient models, Ann. Statist. 27 (1999) 1491–1518.
[17] J. Fan, W. Zhang, Simultaneous confidence bands and hypothesis testing in varying-coefficient models, Scand. J. Statist. 27 (2000) 715–731.
[18] R. Li, H. Liang, Variable selection in semiparametric regression modeling, Ann. Statist. 36 (2008) 261–286.
[19] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman and Hall, London, 1996.
[20] Z. Cai, J. Fan, R. Li, Efficient estimation and inferences for varying-coefficient models, J. Amer. Statist. Assoc. 95 (2000) 888–902.
[21] W. Zhang, S.Y. Lee, X. Song, Local polynomial fitting in semivarying coefficient models, J. Multivariate Anal. 82 (2002) 166–188.
[22] J.-T. Zhang, J. Chen, Statistical inferences for functional data, Ann. Statist. 35 (2007) 1052–1079.
[23] J. Fan, C. Zhang, J. Zhang, Generalized likelihood ratio statistics and Wilks phenomenon, Ann. Statist. 29 (2001) 153–193.
[24] J. Fan, W. Zhang, Generalized likelihood ratio tests for spectral density, Biometrika 91 (2004) 195–209.
[25] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (1982) 405–415.
[26] M.-A. Gruet, A nonparametric calibration analysis, Ann. Statist. 24 (1996) 1474–1492.
[27] P.J. Bickel, M. Rosenblatt, On some global measures of the deviation of density function estimates, Ann. Statist. 1 (1973) 1071–1095.
[28] D. Pollard, Asymptotic for least absolute deviation regression, Econometric Theory 7 (1991) 186–199.
[29] J.A.J. Lee, U-Statistics: Theory and Practice, Marcel Dekker, New York, 1990.